DEVELOPMENT AND APPLICATION OF A CENSUS-BASED REGIONAL RESIDENTIAL GROWTH MODEL FOR BIODIVERSITY RISK ASSESSMENT

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Stephen D. Smith

May 2018

DEVELOPMENT AND APPLICATION OF A CENSUS-BASED REGIONAL RESIDENTIAL GROWTH MODEL FOR BIODIVERSITY RISK ASSESSMENT

Stephen D. Smith, Ph.D.

Cornell University 2018

The USGS National GAP Program is a biodiversity mapping program implemented at the state level via the Cooperative Fish & Wildlife Research Units (CFWRU). The New York CFWRU completed NY-GAP analysis in 2001, providing, for the first time, a statewide vertebrate species distribution dataset. A subsequent regional project, HR-GAP, documented 75% of the State's terrestrial vertebrates as having a significant portion of their range within the Hudson River Valley region (HR). The presence of high biodiversity in conjunction with development pressures was the impetus for efforts to develop a regional residential growth prediction model, based on Block Group (BG) level Census data, with the purpose of identifying biodiversity regions at risk from future residential development.

Initial efforts resulted in a regression model which predicted 77 of the 2,212 total BG in the study area to be prime candidates for a substantial percentage of the predicted new residential growth. These BGs, classified as intensive growth areas (IGA), were intersected with biodiversity data to quantify that 53% of the State's vertebrate species are within and intensive growth BG, as well as 41% of the threatened, endangered, or special concern (TES) species.

Additional model development provided a slight improvement to the predictability of the model while using only digitally available regional data. The second model explained 38% of the variance associated with the identification of IGAs and identified the top 5% of BGs showing substantial increases in residential housing units over the last decade. Of the BGs predicted to be areas of fast growth, 53% and 41% were IGAs as computed from 2000 and 2010 Census data, respectively. Of the IGAs predicted for 2000 and 2010, 16% and 8%, respectively, were also species-rich BGs.

A third modeling effort was undertaken to improve upon the earlier residential housing prediction models based on regression analysis of Census-based BG data and physiographic variables aggregated to the BG level geography. It was hypothesized that increasing the spatial resolution through dasymetric mapping of the BG data would further improve model results and subsequently the identification of biodiversity areas at risk. The model results from the dasymetric mapping did not reveal significant improvement to earlier model results. Investigations of various alternative Census-based datasets yielded similar results.

These efforts to model residential growth at the landscape scale support the hypothesis that the spatial distribution of residential housing growth can be modeled using Census Block Group (BG) level data and other publicly available data to provide a coarse filter for the identification of biodiversity areas at risk from projected residential growth.

BIOGRAPHICAL SKETCH

Stephen D. Smith was born in Waverly, NY in June of 1952. He spent his childhood in Candor and later, Lansing, NY. Stephen earned a B.S. degree in Landscape Architecture from the College of Agriculture and Life Sciences at Cornell University in 1974. In 1987, he completed an M.R.P. degree from Cornell's College of Arts and Sciences under the Employee Degree Program. Stephen began his pursuit of a Ph.D. in 2004, also under the Employee Degree Program. Stephen began his career at Cornell in 1977 as a draftsman at the Resource Information Lab and is now a Senior Extension Associate and Director of the Cornell Institute for Resource Information Sciences. During the last 40 years at Cornell, he has acquired expertise in manual and computerized mapping, Geographic Information Systems (GIS), and Global Navigation Satellite Systems (GNSS) as he assisted countless students and faculty with geospatial projects, theses, and dissertations.

DEDICATED TO

Pamela R. Smith

My wife of 44 years who has seen the best and the worst of me. She has remained by my side with an unwavering love that I cannot possibly deserve. As with everything I do, it has made all the difference.

ACKNOWLEDGMENTS

My sincerest gratitude and appreciation is offered to Milo E. Richmond, my committee chair, who was instrumental in setting me upon this path and in my reaching its end. His support and confidence has never wavered.

My committee members, Patrick Sullivan, Stephen D. DeGloria, and Joe Francis, have provided much insight and guidance, often in the face of my own stubbornness. I highly value their wisdom and friendship over these many years. Their patience has been remarkable.

I would like to thank my co-workers at the Institute for Resource Information Sciences. Thank you Diane Ayers, Susan Hoskins, and Lori Radcliff-Woods for your support and understanding during this effort. I would like to especially thank Magdeline Laba for her inspiration and ever positive attitude.

I would like to express my gratitude to the Department of Natural Resources and the Graduate School for their patience and willingness to accommodate a long-term student.

Acknowledgement of the financial support for the work described herein is given to:

The Hudson River Estuary Program, New York State Department of Environmental Conservation, 21 Putt Corners Road, New Paltz, NY 12561 - Under: Habitat Vulnerability Assessment for the Hudson River Valley Project, Biodiversity and Terrestrial Habitat Project, MOU 5046, Department of Natural Resources, Cornell University.

National GAP Program, United States Geological Survey, Biological Resources Division, 530 South Asbury Street, Suite #1, Moscow, ID 83843. Under: Integration of Socioeconomic Variables into the Hudson River Valley Habitat Vulnerability Assessment Project, Award #1434-HQ-97-RU-01553, Research Work Order 42.

Hudson River Foundation, 17 Battery Place, Suite 915, New York, NY 10004. Under: Habitat Vulnerability Assessment for the Hudson River Valley, Hudson River Research Grant #012/01A.

New York Cooperative Fish and Wildlife Research Unit, College of Agriculture and Life Sciences, Department of Natural Resources, Cornell University, Fernow Hall, Ithaca, NY 14853.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

CHAPTER 1:

INTRODUCTION

Conceptual framework

There have been numerous efforts, over many decades, to predict urban growth. It has been a long-standing goal of planners, builders, developers, and others to develop a method for accurately predicting where and when development will occur. Any attempt to develop a predictive model that, of necessity, engages multiple variables faces difficulty and risk of failure. This possibility appears particularly true if the variables used are both innovative and lacking in precision.

The urban models developed over the years have utilized different predictive techniques. Based on these techniques, four model classes emerge: economic, gravity, optimizing, and hybrid (Foot, 1981).

Linear, econometric, or market models can be traced back to von Thunen and his theory of agricultural ring development influenced by rent dynamics (Alonso , 1964). These models allocate activities to zones or regions, but do not model the spatial interactions between zones or regions.

Working from early concepts of development, migration, and market analysis, models were developed to address spatial interactions not directly included in previous models. From these alterations, the notion of the gravity model emerged.

Gravity models assume that current development impacts new development in a similar manner as bodies of mass interact according to Newton's law of gravitation (Britannica, 2016). That is to say, the intensity of the impact exerted by an existing

development is indirectly proportional to the distance between the potential development area and directly proportional to the size/extent of the existing development area. Hence, a large influencing source at a specific location will have a greater impact on responding factors than a smaller influencing source at the same location. Also, as the distance increases between the location of influencing source and the responding factor, the intensity of the influence decreases.

Alonzo and Lowery undertook the direct application of these two model types, economic and gravity, to residential location (Alonzo, 1964; Lowery, 1964). Many of the recent advances in residential development or growth modeling have their foundations in these two efforts. Over time, these two types have been expanded upon to create optimization, simulation, and hybrid models. A general synopsis of more recent models is included below to provide the larger framework for this dissertation.

Optimization models seek to develop an optimized solution for the landscape based upon land supply and capacity modeling (Moudon & Hubner, 2000). The often mentioned "buildout" model is an optimization model developed around existing zoning and other land use development constraints (Moudon & Hubner, 2000). Simulation models often use similar constraints as "buildout" models but provide the option to modify those constraining inputs, such as re-zoning regions, to investigate effects on the model results (Allen, 2001). These models often include linear and non-linear programming in an iterative process to obtain various results.

Hybrid models are mixtures of any or all three of these model types. They attempt to employ the best modeling method for various activities and spatial interactions to obtain the most realistic results. The development of land use plans to investigate

"what if" scenarios rely largely upon hybrid models using optimization and simulation to develop outcomes based on user changeable inputs (Landis & Zhang, 1998; Kwartler & Bernard, 2001; Berke et al., 2006; Klosterman, 2008).

Many of the models developed have had goals of predicting residential growth at the smallest resolution possible based on the available data. Pursuing this goal results in complex models with more extensive data requirements (Berke et al., 2006). While these requirements may improve the model's effectiveness, they can also impact the cost of model implementation and limit their adoption. For example, if the model requires data that are costly to obtain or develop in a digital form, the model may lose some of its portability to other regions. Alternatively, if a model requires a variable difficult to construct because of required inputs or computational complexity, these hurdles may add to the difficulty of implementation.

My search for a method for the prediction of urbanization began in the mid-1990's with involvement in the NY-GAP Analysis Project (Smith et al., 2001a). During the NY-GAP Analysis Project, Census 1990 data was used to identify the urbanized area as part of the image analysis and habitat mapping efforts. Mapping existing urbanized boundaries provides more accurate urbanized area delineation, reduces the area of analysis required for habitat mapping, and improves the land classification process by reducing classification errors introduced by the urban areas. As work on NY-GAP land cover and habitat mapping progressed, it was apparent that a more general, landscape-level approach to the identification of potential growth areas would be a useful complement to the GAP-scale dataset (Smith et al., 2001a). This realization molded my hypothesis and served as the impetus for the efforts described herein.

My efforts to model residential growth at the landscape scale rest on the hypothesis that the spatial distribution of residential housing growth can be modeled using Census Block Group (BG) level data and other publicly available data to at least provide a coarse filter for the identification of biodiversity areas at risk from projected residential growth.

The development of this landscape model for residential growth prediction began in the fall of 1997 when a Cornell graduate student, Prentiss Sayeweh, in the Department of City and Regional Planning (CRP) approached the Institute for Resource Information Sciences (IRIS) for assistance in identifying a GIS related thesis project. At that time there was interest, at the national level of the GAP Analysis Program, to add demographic data into the individual State GAP projects. Because of this interest, assistance was provided to the graduate student in developing a Master of Regional Planning thesis project with support provided by IRIS. The focus of Mr. Sayeweh's thesis project was the development of a habitat vulnerability assessment methodology based on projected population growth and existing developable land. The study area for the project, Putnam County, New York, is within the Hudson River Valley region. (Sayeweh, 1998).

I submitted a proposal to the New York State Department of Environmental Conservation (NYSDEC) to fund a Habitat Vulnerability Assessment in the Hudson River Valley based on Mr. Sayeweh's early work. Unfortunately, it was many months after Mr. Sayeweh's graduation and departure from Cornell that NYSDEC funded the project. Ultimately, in 2000, the NYSDEC Hudson River Estuary Program (HREP)

recognized the value of predicting housing development and its potential negative impact on biodiversity and approved funding for the assessment of habitat vulnerability in the Hudson River Valley region.

Study area

With funding and project approval in hand and the heightened interest of the HREP, the study area expanded to the entire Hudson River Valley (HRV) region. The HRV ten-county region includes all of the following counties: Albany, Columbia, Dutchess, Greene, Orange, Putnam, Rensselaer, Rockland, Ulster, and Westchester (Figure 1.1). Contained within a rectangle with corner coordinates of 74^0 50' W, 40^0 50' N (lower left) and 73^0 10' W, 43^0 0' N (upper right), the study area extends from Albany County to New York City.

This ten-county region, with the addition of adjacent New York City, includes approximately half of New York State's human population while at the same time providing habitat for hundreds of migratory and resident species of wildlife (Smith et al., 2001a). This region is facing a period of re-industrialization, residential development, and constant development pressure emanating from New York City northward and, to a lesser degree, southward from the city of Albany (Hu, 2000). The region possesses substantial political power and associated funding potential. This ten-county region, bisected by the lower reach of the Hudson River, has a high degree of environmental awareness and activism. This environmental consciousness may perhaps have roots in the naturalistic art identified as the Hudson River School in the

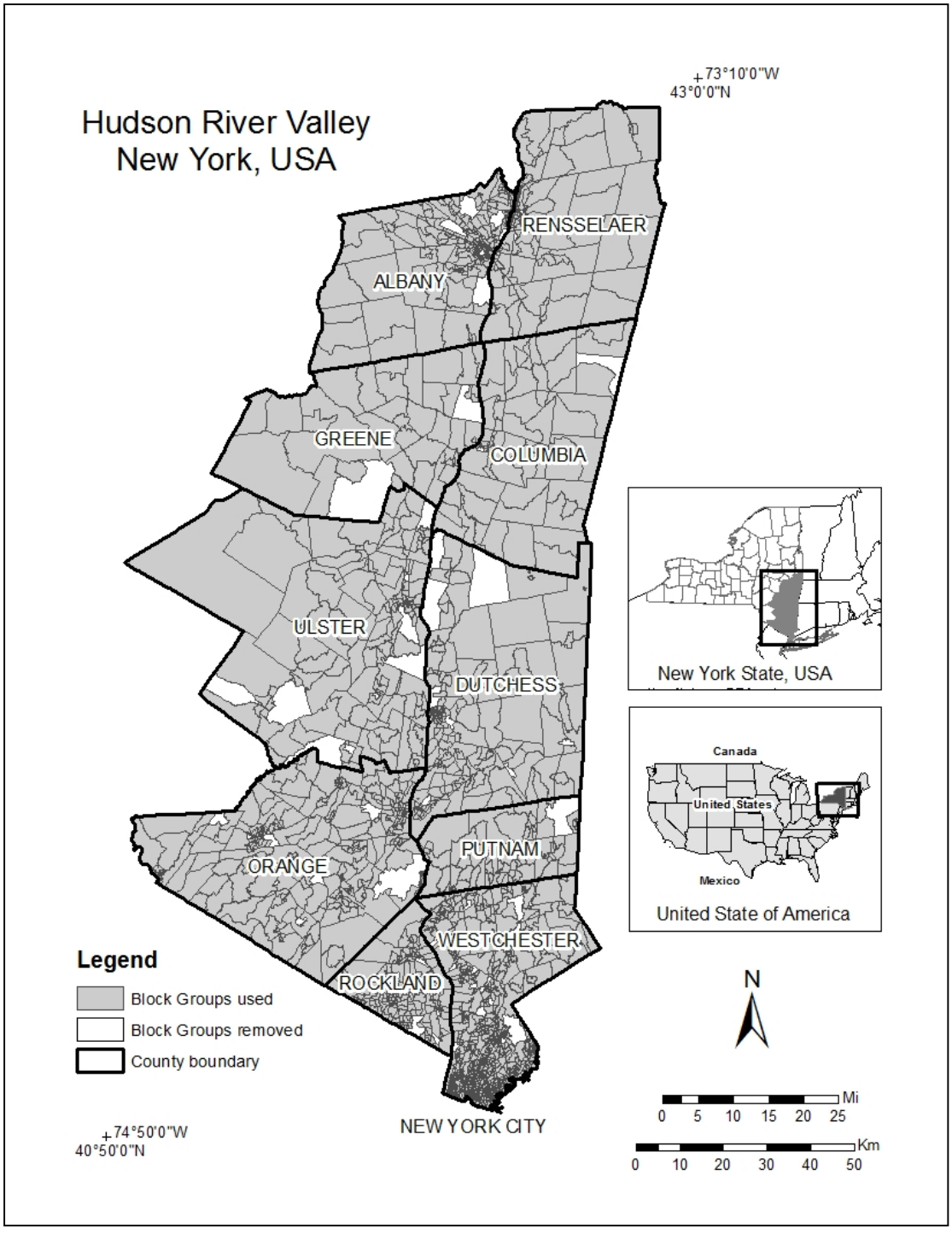


Figure 1.1. Ten-county region with the 2,111 Block Groups used in the analysis and the 101 omitted Block Groups that represented group quarters and institutional housing. Group quarters and institutional housing increases are subject to institutional and governmental forces outside the scope of this study.

mid-1800's (Metropolitan Museum of Art, 2017). From wherever this consciousness grew, it is evident in recent developments such as the establishment of Scenic Hudson, an agrarian environmental group, in 1963 (Scenic Hudson, 2017); Riverkeeper, a watchdog organization monitoring the Hudson River and associated tributaries, also in 1963 (Riverkeeper, 2017); Hudson River Clearwater Sloop, a grassroots environmental education and action group, in 1969 (Clearwater, 2017); and the Hudson River Estuary Program, a NYSDEC program providing grants, education, resource conservation and community assistance, in 1987 (Hudson River Estuary Program, 1987).

The valley is also critical from a biological perspective. The HRV provides habitat for 69% of all NY resident amphibian species (25 species) and 58% of all NY resident reptile species (28 species) found in New York (Smith et al., 2001b). Habitat for 87% of all NY resident breeding bird species (214 species) and 90% of all NY resident mammal species (57 species) is also provided (Smith et al., 2001b). Among NY resident terrestrial vertebrates, 75% have all or a significant portion of their range within the Hudson River Valley study area (Smith et al., 2001b). This high concentration of less mobile vertebrate species and the significant contributions from all four vertebrate groups strengthens the importance of this ten county region as a study area.

The region is diverse in both natural and anthropogenic characteristics as shown in Figures 1.2 and 1.3, respectively. For selected anthropogenic characteristics of each county as a percentage of the total study area see Figure 1.2.

A substantial county-to-county diversity is evident with five counties (Albany, Dutchess, Orange, Putnam, and Rensselaer) having internal percentage values that vary relatively little, less than 5%, between anthropogenic characteristics while the remaining five counties provide the intra-county diversity (Figure 1.2). Refer to Chapter 4 for a specific discussion of the county-to-county variability of the anthropogenic characteristics presented in Figure 1.2.

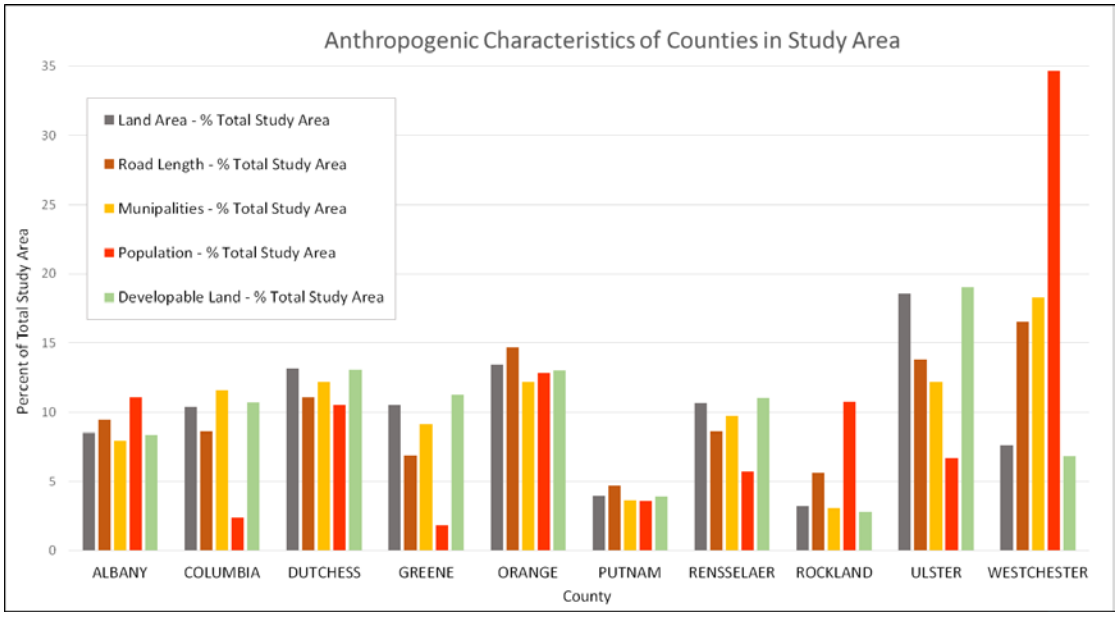


Figure 1.2. Selected anthropogenic characteristics of each county in the study area as a percentage of the total study area characteristics.

The physical diversity of the region is graphically displayed by the substantial variation of percentages for selected natural characteristics within each county as well as between the counties (Figure 1.3). To provide a shared context, land area distributions shown in Figure 1.3 are the same as those in Figure 1.2. Refer to Chapter

4 for a specific discussion of the county-to-county variability of the natural characteristics presented in Figure 1.3.

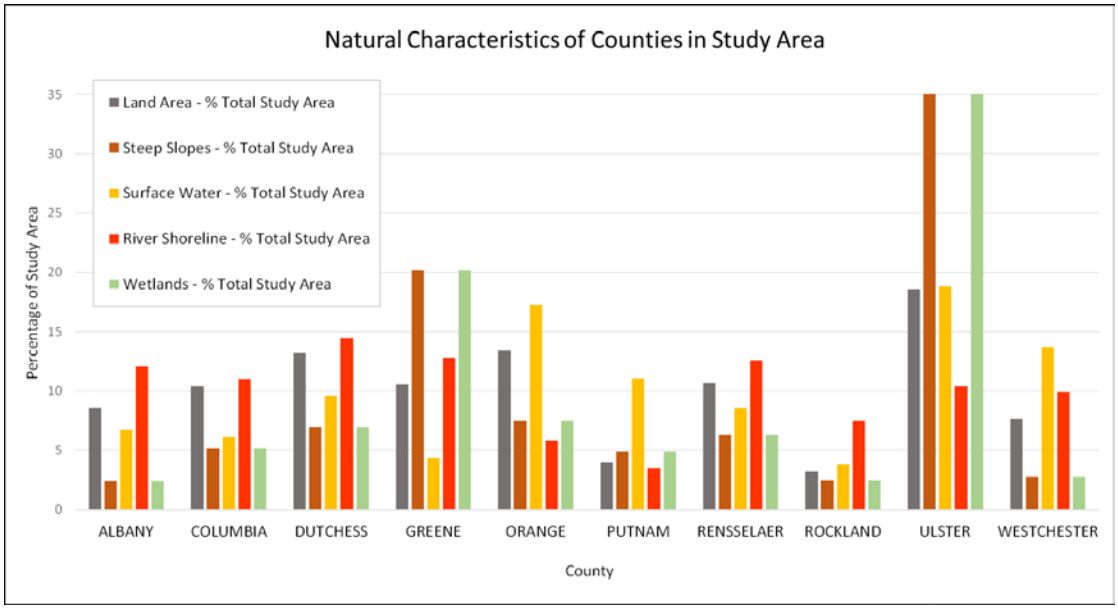


Figure 1.3. Selected natural characteristics of each county in the study area as a percentage of total study area characteristics.

Building the research model

Based on the desire to develop an urbanization model as a complement to GAP and based on the amount of effort expended to compile the datasets required for the modeling efforts in Putnam County, one of New York's smallest counties, three goals were set to guide model development. These goals are: 1) provide a coarse filter for the regional identification of potential high residential growth areas; 2) utilize existing and publically available digital data; 3) and extend portability to other areas of New York and other states.

With these three goals as a guiding framework, Dr. Warren Brown, at the Cornell Institute for Social and Economic Research, was approached about assisting in developing a new, less labor-intensive approach that would only require readily available digital data. The initial version of this growth allocation model included census-based social ecology data with physiographic and population density information. The social ecology sector composition includes factors based on race, economic class, urban status, lifestyle, and mobility ratings. This approach was new and especially attractive in that it was intuitively logical and could be easily implemented anywhere in the United States. In addition to the social ecology factors, the model includes counts of housing units built in the prior decade, labor market areas, road network density, and proximity to selected population centers.

Initial runs of the nine variable model showed that the five census-based social ecology factors lacked statistical significance concerning the model's results. Although the remaining four factors are significant, other independent variables were needed. The investigation into Birch's Neighborhood Stages (Birch, 1971) concept seemed promising and proved to be a suitable replacement for the social ecology sector (Yang et al., 2001). The finalized model included a housing stage factor based on Birch's six neighborhood stages. The addition of Birch's Neighborhood Stages (BNS) to the model increased the predictive power to 37% of the variability and made possible its identification of 77 of the 2212 BGs as "potential intensive growth areas" for residential development. The potential intensive growth areas were intersected with the NY-GAP data (Smith et al., 2001a) aggregated to the BG level, to identify high species richness areas coincident with potentially high residential growth areas.

The results from this model support the hypothesis that residential growth can be modeled using Census data and that the results from such modeling can be used to efficiently identify biodiversity risk areas, thereby enabling the targeting of education and mitigation efforts. However, the need to use other publicly available data that were not in digital form was not consistent with the goal to confine inputs to existing digital data. An adjustment to the model was necessary to bring it back into line with my stated goals.

Adjusting the model

This model worked well for the study area, the ten counties of the Hudson River Valley: Albany, Columbia, Dutchess, Greene, Orange, Putnam, Rensselaer, Rockland, Ulster, and Westchester. However, shortcomings in both its development and testing due to a lack of data from the Census 2000 survey was a concern. Model development using 1990 CENSUS data and 1995 CENSUS estimates could only be tested against these estimates. The CENSUS 2000 data were needed to confirm the model, but the release of the CENSUS 2000 data did not occur within the funding period of the HREP. As a consequence, I was not able to finalize the model development without further funding.

Additional funding to bridge this lapse in financial support was eventually gained from the USGS National GAP Program via the New York Cooperative Fish & Wildlife Research Unit, Department of Natural Resources, Cornell University, to pursue the goal of demographic data integration with GAP results. Because my ultimate goal was to develop a growth allocation model that, when implemented,

would identify potential growth areas coincident with areas of high biological diversity, the National GAP Program became very interested in my research results. This funding was instrumental in developing the sociological aspects of the model. In the summer of 2000, a paper describing the results was presented at the National GAP meeting at San Antonio, Texas.

As with many research projects, I began with a concept and some seed money and, as the concept developed and grew, other funding sources became interested. The Hudson River Foundation (HRF) saw the earlier work, funded by HREP and USGS, and responded favorably to a proposal to continue, expand, and refine the methodology. With the additional support provided by the HRF, I was able to bring to a conclusion the development of the growth allocation model based on CENSUS 2000 data and implementation of the model in the assessment of risk to areas of high biodiversity in the Hudson River Valley.

This approach was new and innovative. The development of the model has not been without risk, as any research project is, but has also held promise. Due to the HREP and the USGS initial funding associated with habitat vulnerability, an additional $220,000 of funding was leveraged over subsequent years to address the issue of Hudson River Valley Habitat Vulnerability.

The efforts described above are detailed in Chapter 2: Census-based Residential Growth Model for Habitat Vulnerability Assessment.

Eliminating Birch's Neighborhood Stages

Chapter 3: Modeling Residential Growth and Habitat Vulnerability on a Regional Scale, is focused on the next iteration of model development. Although the Birch's Neighborhood Stages (BNS) factor provided an increase in predictive power, the amount of local information needed to develop the variable was substantial and counter to the goal of the minimizing local non-digital data requirements. To get the model back on the track of utilizing only available digital data, an alternative variable to replace BNS was sought.

This chapter focuses on the hypothesis that modeling the spatial distribution of residential housing growth with Census Block Group (BG) level data can be accomplished without requiring non-digital local data, specifically, the data needed for the calculation of the BNS factor while providing similar predictive power for a suitable coarse filter biodiversity risk application.

The resulting model eliminated the BNS variable, used three of the previous variables, and added two new variables. The five variables used were: 1) current housing unit density; 2) current road density; 3) previous decade's housing unit change; 4) housing unit density gradient derived from adjacent BGs; and 5) population potential index. This model explained 38% of the variance associated with the identification of intensive residential growth areas and enabled the identification of the top 5% of block groups showing increases in residential housing units over the last decade. This model improved upon the previous model in that it predicted housing unit growth for each BG directly. The earlier residential growth model was a growth

allocation model which distributed forecasted housing development among the BGs based on predictions of shared distribution. The prediction of the increased number of housing units provides easily understandable results without the need to rely upon a regional prediction.

This model's easier data compilation with similar predictive power provides a substantial improvement. In addition to the traditional statistical testing available, error matrix information (Congalton & Green, 1999; Lillesand & Kiefer, 2000; Campbell & Wynne, 2011) was used to measure the accuracy of the model results. An error matrix was implemented because the availability of the Census 2010 data provides observed data against which the predicted model results could be directly evaluated. Also, the error matrix lends itself well to measuring and visually representing the level of accuracy of predicting residential housing growth areas.

Having stepped away from using the Birch's Neighborhood Stages variable, the model was able to obtain similar results in predicting residential growth areas as measured at the BG level. Additionally, these identified areas were shown to overlap more biodiversity risk areas than the previous model. Since my ultimate goal is biodiversity conservation, this increased coincidence is critically important.

The success of this refined modeling effort supports the hypothesis that the non-local data were not necessary to maintain the predictability of growth areas and maintain the applicability of the model to biodiversity conservation.

These encouraging results led to the idea that the way to improve the model's predictive power was to increase the spatial resolution of the variables which dasymetric mapping could accomplish.

Dasymetric mapping, as used in this context, is the process of remapping data aggregated or mapped within a specific geography in an attempt to parse that data into higher resolution geographies that are subsets of the original, using ancillary data. For example, given a square mile town comprised of 45% forest, 45% urban, and 10% water and with a total population of 10,000 persons, a population density of 10,000 persons per square mile can be assigned to the entire town. This town-wide density is not helpful if there is a need to identify areas where the town should increase the availability of human services. However, it is intuitive that the real population density of the urban area within the town is much higher than the forest area and that the population density of the water area is zero. Dasymetric mapping would take the areas of forest, urban, and water within the town and attempt to allocate appropriate proportions of the town's total population to those areas based on ancillary data such as water and sewer availability.

Dasymetric mapping of the Census data entails the apportionment of Census data to subset regions of the BG based on higher resolution physiographic characteristics of the BG. In a limited sense, some dasymetric mapping was done when open water areas were removed prior to calculating housing densities and developable land in the first and second models, respectively.

Investigating dasymetric mapping

I hypothesize, based on the earlier model successes, that the spatial accuracy of the existing residential growth model can be improved using dasymetric methods to

distribute BG data to spatial subsets while maintaining the effectiveness of the model to identify at-risk biodiversity areas.

Chapter 4: Impact of Dasymetric Mapping on a Census-based Regional Scale Residential Growth Model presents the investigation of dasymetric mapping effects on the model. An earlier study (Holt et al., 2004) showed that when applied in a regional scale project in the Atlanta area of Georgia, dasymetric mapping improved the assignment of Tract-level data across decadal surveys with the assistance of land cover data.

I have undertaken the effort to implement the dasymetric mapping of Block Group data to smaller polygons defined by the intersection of Census BG geography and National Land Cover Database (NLCD) polygons (Mennis & Hultgren, 2006; Homer et al., 2007), independent of tax parcels.

The model for this effort uses the NLCD 2001 (USGS, 2014) land cover data to distribute the Census 2000 BG housing units (HU) onto the landscape. The NLCD 2001 land cover data was selected because of efforts to produce reliable updates over time. For this same reason, this effort also used the GeoLytics Census product (GeoLytics, 2000a, GeoLytics, 2000b) which enables analysis across the decadal surveys and was used in the earlier model.

The first attempt at dasymetric mapping involved using the NLCD data to subdivide the BG by land cover class. The methods investigated used the NLCD land cover dataset at multiple resolutions (30-meter, 90-meter, 180-meter, and 270-meter) to acquire a housing unit range to associate with five land cover types generalized from the original 21 classes of the NLCD 2001 dataset. These BG land cover regions

were further dissected using a restricted layer identifying areas from which development was excluded for physical or legal reasons. Regardless of spatial resolution, the five-variables applied to the BG land cover subsets as defined failed to achieve the level of accuracy demonstrated by the full BG analysis.

Further investigation showed a substantial misalignment between the Census 2000 BG boundaries and the NLCD 2001 land cover data set. A review of reliable base map information and recent aerial imagery showed that many of the Census 2000 boundaries were misaligned, but not in a manner that could be easily corrected by shifting or similar manipulation.

To address this unexpected misalignment issue, two new Census derived datasets were acquired. The NHGIS Time Series dataset (Manson et al., 2017) and the GeoLytics 2000 in 2010 BG Boundaries dataset (GeoLytics, 2010). These datasets provide Census 2000 BG data in Census 2010 boundaries using different methods to accomplish the spatial translation. The model was tested against each of these datasets, and a discussion of the results is presented in Chapter 4.

The results presented in Chapter 4 support the null hypothesis. The introduction of dasymetric mapping techniques not only failed to improve the existing model, but instead altered the model in such a way as to make it unacceptable for growth predictions and subsequent detection of at-risk biodiversity areas.

References

Allen, E. (2001). INDEX: Software for Community Indicators. In R. K. Brail & E. Klosterman (Eds.), *Planning Support Systems* (pp 229-261). Redlands, CA: ESRI Press.

Alonso, W. (1964). *Location & Land Use - Toward a General Theory of Land Rent*. Cambridge, MA: Harvard Press.

Berke, P.R., Godschlk, D.R., Kaiser, E.J. & Rodriguez, D.A. (2006). *Urban Land Use Planning*, 5th Ed., Urbana, IL: University of Illinois Press.

Birch, D.L. (1971). Toward a Stage Theory of Urban Growth. *Journal of American Institute of Planners*, 37(2), 78-87.

Britannica (2016). In *Encyclopedia Britannica* online. Retrieved from Encyclopedia https://www.britannica.com/science/Newtons-law-of-gravitation

Campbell, J.B. & Wynne, R.H. (2011). *Introduction to Remote Sensing*. New York, NY: The Guilford Press.

Clearwater (2017). Hudson River Clearwater Sloop. Retrieved from http://www.clearwater.org/

Congalton, R.G. & Green, K. (1999). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, New York, NY: Lewis Publishers.

Foot, D. (1981). *Operational Urban Models an Introduction*. Bungay, UK: Chaucer Press.

GeoLytics. (2000a). *Census 2010 Summary File 1 in 2000 Boundaries*, Geolytics CD-ROM.

GeoLytics. (2000b). *Census 1990 Summary File 1 in 2000 Boundaries*, Geolytics CD-ROM.

GeoLytics. (2010). Census Data 2000 in 2010 Boundaries, GeoLytics, Inc., Retrieved from http://www.geolytics.com/USCensus,Normalized-Data,Categories.asp

Holt, J.B., Lo, C.P., & Hodler, T.W. (2004). Dasymetric Estimation of Population Density and Areal Interpolation of Census Data. *Cartography and Geographic Information Science*, 31(2), 103–121. https://doi.org/10.1559/1523040041649407

Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C.,…Wickham, J. (2007). Completion of the 2001 National Land Cover Database for the Conterminous United States. *Photogrammetric Engineering and Remote Sensing*, 73(4), 337-341.

Hu, W. (2000, June 7). Industrial Growth Threatens Scenic Hudson River Valley, Group Warns. *New York Times*. Retrieved from http://www.nytimes.com /2000/06/27/ nyregion/ industrial-growth-threatens-scenic-hudson-river-valley-group-warns.html

Hudson River Estuary Program (1987). Hudson River Estuary Management Act. *NYS Environmental Conservation Law*, 11-0306. Retrieved from http://www.dec.ny.gov/lands/43251.html

Klosterman, R.E. (2008). A New Tool for a New Planning: The What if? Planning Support System. In R.K. Brail (Ed), *Planning Support Systems for Cities and Regions*, (pp. 85-99). Cambridge, MA: Lincoln Institute of Land Policy.

Kwartler, M. & Bernard, R.N. (2001). CommunityViz: An Integrated Planning Support System. In R. K. Brail & Klosterman, E. (Eds.), *Planning Support Systems,* (pp. 285-308). Redlands, CA: ESRI Press.

Landis, J. & Zhang M. (1998). The second generation of the California urban futures model. Part 2: Specification and calibration results of the land-use change submodel. In *Environment and Planning, B. Planning and Design*, 25(6), 795-824

Lillesand, T.M., & Kiefer R.W. (2000) *Remote Sensing and Image Interpretation,* New Jersey: John Wiley Press.

Lowery, I.S. (1964). *A Model of Metropolis, Memorandum RM-4035-RC*, Santa Monica, CA: The Rand Corporation.

Manson S., Schroeder J., Van Riper D., & Ruggles S. (2017). *IPUMS National Historical Geographic Information System: Version 12.0 [Database]*. Minneapolis, MN: University of Minnesota. http://doi.org/10.18128/D050.V12.0

Mennis, J., & Hultgren, T. (2006). "Intelligent" Dasymetric Mapping and Its Comparison to Other Areal Interpolation Techniques. *Cartogis.Org,* (215), 1–9. Retrieved from http://www.cartogis.org/publications/autocarto-2006/mennishultgren.pdf

Metropolitan Museum of Art. (2017). *Heilbrunn Timeline of Art History*. Retrieved from https://www.metmuseum.org/toah/hd/hurs/hd_hurs.htm2

Moudon, A. & Hubner, M. (2000). Elements of a General Framework for Land Supply and Capacity Monitoring. In A. Moudon & M. Hubner (Eds), *Monitoring land supply with geographic information systems: Theory, practice, and parcel-based approaches* (pp. 41-64). New York, NY: John Wiley & Sons.

Riverkeeper (2017). https://www.riverkeeper.org/riverkeeper-mission/our-story/

Sayweh, P.W. (1998). *Combining a Growth Allocation Model and GIS Results to Assess Suitable Biodiversity Management Areas: A Prototype for Hudson Valley Biodiversity Program in New York State* (Unpublished MRP thesis), Cornell University Graduate School, Ithaca, NY.

Scenic Hudson (2017). https://www.scenichudson.org/about/history

Smith, C.R., DeGloria S.D., Richmond M.E., Gregory S.K., Laba M., Smith S.D.,…Weber J.T. (2001a). *A Gap Analysis of New York NY-GAP. Final Report and Data*. U.S. Geological Survey, Biological Resources Division, National Gap Analysis Office, Moscow, ID. 2-Disc CD set (issued October 2001).

Smith, C.R., DeGloria S.D., Richmond M.E., Gregory S.K., Laba M., Smith S.D.,… Hill E.A. (2001b). *An Application of Gap Analysis Procedures to Facilitate Planning for Biodiversity Conservation in the Hudson River Valley, Final Report, Part 1: Gap Analysis of the Hudson River Valley and Part 2: Atlas of Predicted Ranges for Terrestrial Vertebrates in the Hudson River Valley*. New York Cooperative Fish and Wildlife Research Unit. Ithaca, NY:Department of Natural Resources, Cornell University.

USGS (2014). *NLCD 2001,* USGS National Land Cover Dataset Downloadable Data Collection, January, 2014, https://viewer.nationalmap.gov

Yang, Y., Brown W., & Smith S. (2001). *Using Neighborhood Growth Stage Model to Analyze New Residential Development in Ten Counties in Hudson River Valley Area, New York State.* Population & Development Program Working Papers. Ithaca, NY: Program in Applied Demographics, Cornell University.

CHAPTER 2:

CENSUS-BASED RESIDENTIAL GROWTH MODEL FOR HABITAT VULNERABILITY ASSESSMENT

Abstract

This paper describes the efforts undertaken and the results obtained in a study to assess habitat vulnerability arising from residential housing growth, for the Hudson River Valley (HRV). The goals of this study were: 1. Create a residential development model derived from US Bureau of the Census data to predict areas where future residential development is likely to occur, and 2. Highlight those areas where such development is coincident with species richness.

This study focused on a ten-county region of the Hudson River Valley of New York State. The counties of Albany, Columbia, Dutchess, Greene, Orange, Putnam, Rensselaer, Rockland, Ulster, and Westchester define the study area.

Using Census-based data aggregated to the Block Group (BG) geography, 3.5% of the 2,212 BGs involved (77) were classified as prime candidates to receive major shares of the new housing projected to be built within the 2000 to 2010 period. Five independent variables drive the model. These independent variables are neighborhood stage of development; number of housing units built in the prior decade; regional labor market area; density of local road network; and proximity to centers of population. The model explains 37% of the variation in the level of residential development.

After determining the BGs likely to receive the predicted new residential housing, BGs containing areas of high species richness were identified. For each vertebrate

species group, a BG weighted count was calculated. Weighted counts for each vertebrate group's threatened, endangered, and sensitive (TES) species were also calculated. These calculations enabled the mapping of expected species richness relative to the total expected species richness for each group.

The intersection of these species richness areas and the prime BG candidates for residential development highlighted the species richness areas most vulnerable to potential future residential development. This identification of vulnerable areas permits the prioritization of potential residential development BGs with high biodiversity into areas for concern or governmental action.

All of the data required by the model were acquired or derived from readily available public data at a spatial resolution appropriate for regional studies. The model requires no detailed zoning data, thus increasing the applicability of this model to other regions in the New York or other states. As a means to identify areas of potential concern, the model is a practical and useful tool for county and regional planning.

Introduction

The ten counties of the Hudson River Valley corridor support almost a third of the human population of New York State north of New York City while providing habitat for hundreds of migratory and resident species of wildlife. The NY Gap Analysis Project (NY-GAP) identified the vertebrate species found within New York State and found that the Hudson River Valley possessed habitat supporting over 80% of the State's terrestrial vertebrate species (C. Smith et al., 2001a). A regional gap analysis

project focused on the Hudson River Valley (HRV) has confirmed these results (C. Smith et al., 2001b).

GAP analysis is geographic methodology developed to provide a coarse filter for protecting biodiversity (Scott et al., 1993). GAP has as its goal, the identification of habitat and species diversity areas that exist outside the boundary of any legal mechanism for habitat preservation. The National GAP Analysis Project was an undertaking by the US Fish and Wildlife Service to develop a national dataset by supporting the completion of individual statewide GAP projects through the National Cooperative Fish and Wildlife Research Unit Progam structure located at nearly all state land grant colleges and universities. The New York State Cooperative Fish and Wildlife Research Unit, in the Department of Natural Resources at Cornell University was responsible for the completion of the first New York State GAP dataset (C. Smith et al., 2001a).

The Hudson River GAP project (HRV-GAP) identified the extent to which the HRV contributes to statewide vertebrate biodiversity, as measured by vegetation associations and terrestrial vertebrate occurrences, in coarse-filter and fine-filter categories (C. Smith et al., 2001b). HRV-GAP found that 25 of the 31 coarse-filter elements of biodiversity (i.e. super-alliances) identified in the State are found in the HRV. Among those super-alliances, the HRV-GAP states that the HRV has proportionately "more Sugar Maple-Mesic Forest, Oak Forest, and Appalachian Oak-Pine Forest than found elsewhere in New York State" (C. Smith et al., 2001b Part1 pp.6.1) embedded predominantly in an urban/suburban matrix. Coarse-filter vegetative diversity is not limited to rural counties. The primarily rural region of Ulster Co. and

the more urbanized region of Westchester Co., each contain more than twenty vegetative super-alliances within their boundaries.

For all terrestrial vertebrates combined, 86% (308 species) are documented to occur within the HRV and many of New York's terrestrial vertebrates have a significant proportion or all of their entire ranges within the HRV study area (C. Smith et al., 2001b). HRV-GAP report notes that the variety of herpetofauna represented within the HRV offers conservation opportunities for amphibian and reptile biodiversity not available elsewhere in the State (C. Smith et al., 2001b Part 1 pp.6.2).

The above data reveal the biological importance of the HRV to the overall ecology of New York State. This region plays an equally important role in the economic stability and cultural heritage of New York State. It is a region under constant development pressure emanating primarily from New York City. The pressures for commercial, industrial, and residential growth continue to impact the biodiversity of the area. These development impacts also impact the ecology of the Hudson River and its tributaries. These impacts may be direct and indirect. Direct impacts, such as the removal of wetlands or species, are often more evident. Indirect impacts, such as the fragmentation of habitat or increased use of public facilities, may be subtle but may result in more sustained damage. The ability to accurately project the location of residential development will empower decision makers. Knowing areas of future growth enables the identification of ecological and cultural areas potentially threatened. This knowledge can inform mitigating actions to protect wildlife and fish habitats, diversity, and significant cultural sites.

The goals of this study were: 1) Create a residential development model derived from US Bureau of the Census data to predict areas where future residential development is likely to occur, and 2) Highlight those areas where such development is coincident with species richness.

There were three objectives for this study. The first objective was to create a model to accurately predict where, within the region, residential development is likely to occur. The second objective was to parameterize this model using uniform and nationally available digital data. The third objective was to identify the area of coincidence between the BGs identified by the model as probable regions of substantial housing growth and BGs identified as species richness areas by the HRV-GAP project.

The basis for this effort is the hypothesis that the spatial distribution of residential housing growth can be modeled using Census BG level data and other publicly available data to provide a coarse filter for the identification of biodiversity areas at risk from projected residential growth.

Study area

The Habitat Vulnerability study area covers ten counties in the HRV of New York State. These counties include Albany, Columbia, Dutchess, Greene, Orange, Putnam, Rensselaer, Rockland, Ulster, and Westchester (Figure. 2.1). The biological importance of this region has been documented by the NY-GAP report (C. Smith et al., 2001a). A subsequent, regional gap analysis project focused on the Hudson River Valley has confirmed this importance (C. Smith et al., 2001b).

As described in the HRV-GAP report, “Among terrestrial vertebrates (fine-filter biodiversity elements), 85% (28 species) of NY's total amphibian species, 73% (27 species) of NY's total reptile species, 87% (199 species) of NY's total breeding bird species, and 92% (54 species) of NY's total mammal species can be found in the HRV (C. Smith et al., 2001b Part1, pp iii). Among terrestrial vertebrates, some have all or a significant proportion of their entire NY ranges within the HRV study area: Marbled Salamander, Bog Turtle, Northern Fence Lizard, Eastern Worm Snake, Copperhead, Black Vulture, Blue Grosbeak, and New England Cottontail. In general, the HRV offers opportunities not found elsewhere in NY for the conservation of amphibian and reptile biodiversity because of the variety of herpetofauna represented there, especially turtles.

The biological importance of the HRV to the ecology of New York State is shown by the above statistics. This region plays an equally important role in the State’s cultural heritage and economic stability.

This concentration of biodiversity occurs in a region under constant development pressure mainly emanating from New York City. Additionally, the Hudson River Valley may well be facing a period of joint re-industrialization and residential development. This re-industrialization and residential growth will impact the biodiversity of the region and thereby the ecology of the Hudson River. These impacts may be as direct as the removal of wetlands, plant species, and the dislocation of animal species or as indirect as the reduction of wildlife habitat below a sustainable

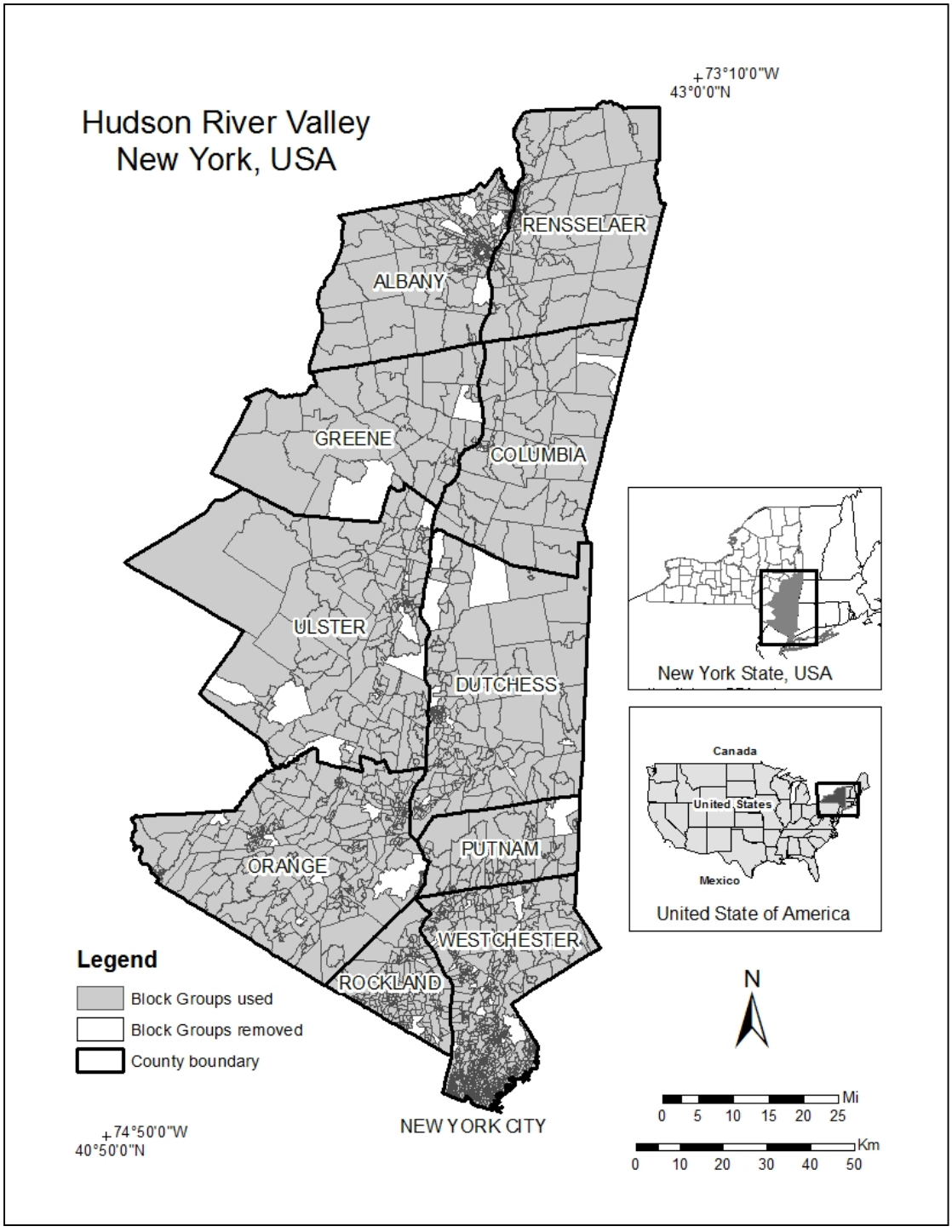


Figure 2.1. Ten-county region with the 2,111 Block Groups used in the analysis and the 101 omitted Block Groups that represented group quarters and institutional housing. Group quarters and institutional housing increases are subject to institutional and governmental forces outside the scope of this study.

level, increased stormwater runoff into tributaries, increased public use of already challenged facilities, and changes to the cultural composition of a neighborhood. Few things have as serious an impact on the ecology of place as human development. The ability to accurately predict the loci of that development and subsequently identify the ecologically and culturally sensitive areas susceptible to the resulting impacts will empower decision makers with the knowledge necessary to take actions required to minimize impacts and maintain wildlife and fish habitat, biological diversity, and regionally significant historical/cultural sites.

Methods

Predicting potential areas of intensive residential development

The unit of analysis for the modeling of residential development is the Census BG geography. The decennial census of population uses BGs for persons and housing data tabulation to avoid confidentiality concerns that arise with the tabulation of the same variables at the Block level. Tracts, Block Groups, and Blocks are the geography used for reporting Census tabulations (Figure 2.2). A statistical unit of census geography, the BG includes approximately 600 housing counts and as many as 3,000 persons, although the optimum size is 1500 people (USCB, 2002). The 2000 Census provided data for the 2,212 BGs within the Hudson River Valley region. Long-form Census questionnaires from a sample of households provided data that were tabulated and summarized by the Census Bureau for each BG (USCB, 1994; USCB, 1999; USCB, 2000).

To determine the level of residential development among Block Groups I used housing unit (HU) data from the Census 2000 to identify the year built. Census 1990 provided data for other variables needed to explain residential development variations among BGs.

The original model included a set of four variables: 1) number of housing units built in the prior decade; 2) regional labor market area; 3) density of the local road network; and 4) proximity to centers of population. To these original four variables were added five additional variables derived from Census data and identified as social ecology factors: 1) socio-economic status; 2) family life cycle; 3) mobility; 4) race; and 5) urban/rural.

Initial model runs with all nine variables showed that the five census-based social ecology factors lacked statistical significance concerning the model's results. Although the remaining four factors are significant, other possible independent variables were needed. An investigation into Birch's Neighborhood Stages concept (Yang et al., 2001) seemed promising and proved to be a suitable replacement for the social ecology sector.

The best model explaining residential development variations had five independent variables: number of housing units built in the prior decade, the density of local road network, neighborhood stage of development, regional labor market area, and proximity to centers of population. A binomial model was applied to these five variables because a large number of BGs had no new housing construction and regional housing forecasts existed which could be used to allocate housing.

Dependent variable: new housing units

I selected HUs built between January 1, 1990, and April 1, 2000, as the count of new HUs. The decennial census asks residents the year of construction for the structure containing their HU. Examples of housing unit types, as counted by the census, include single-family houses, duplexes, apartments, and mobile homes. In the case of vacant housing, enumerators try to make an informed determination of the HU age based on information from neighbors or building superintendent where applicable. The long form questionnaire, administered to a sample of households (approximately 1-in-6), includes the year built question which is used to generate BG

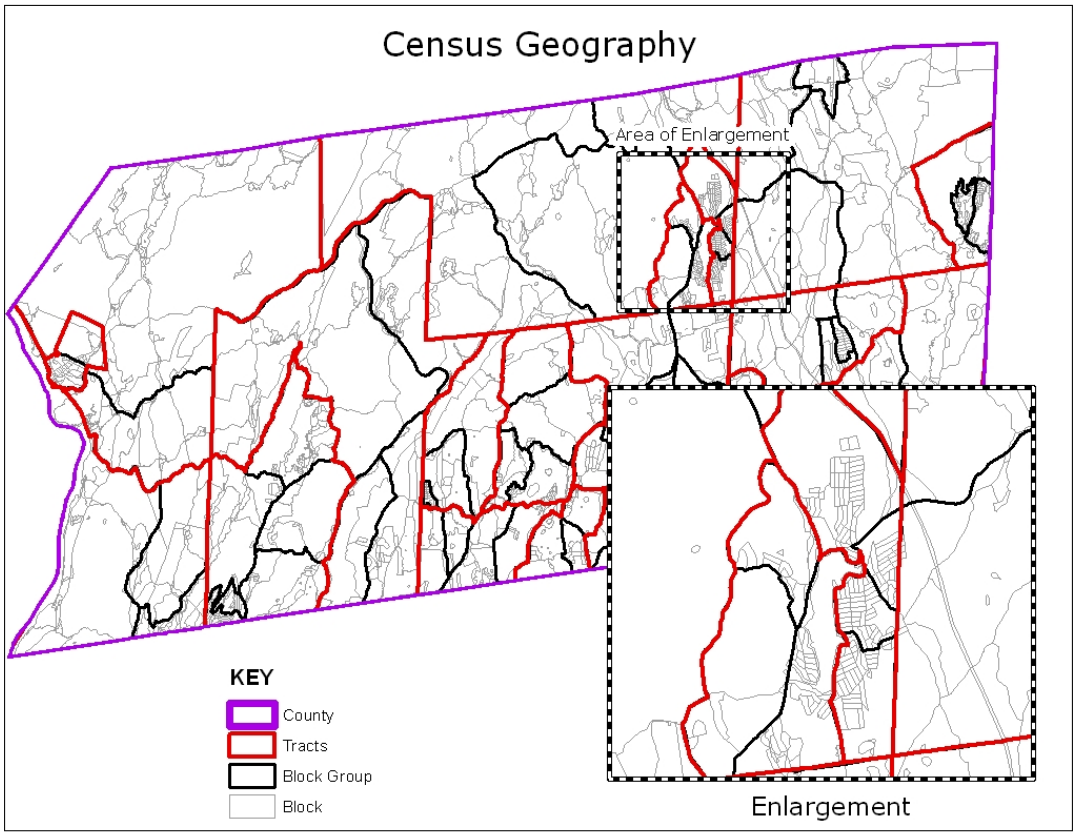


Figure 2.2. Census geography of County, Tract, Block Group, and Block.

estimates. The Census Bureau provides summaries of these sample data for various geographies from BGs up to county level.

A relatively small proportion of the BGs contain most of the HRV's residential development. Eleven percent of the BGs account for half the new housing units. One-in-five of all new units occupy 3% of the BGs. The number of new HUs within a BG in the HRV varies from a high of 637 to zero. More than 40 percent of the BGs in the region have ten or less new housing units (Figure 2.3).

Rather than analyze the variation in the number of new HUs by BG, I converted the dependent variable to a percent of new HUs. The primary reasons for the conversion are the assumptions that regional housing markets explain the aggregate level of residential development and neighborhood models accomplish the allocation of those new units. Explaining the total of new HUs within the study area was not my objective. I already had a forecast (multi-county) of residential development, so my task was to predict which BGs were likely to receive those new HUs.

Independent variable: neighborhood stage of development

BG data from the 1990 Census provided housing characteristics and neighborhood development information, enabling the calculation of neighborhood stage classes such as single-family subdivision, built-up, structure conversion, downgrading and renewal (Hoover & Vernon, 1959, pp 190-207). I adopted Birch's (Birch, 1971) classification of neighborhood growth stages for this project (Figure 2.4). The six stages categorize neighborhood development into quasi-sequential classes, all of which may not be experienced by all BGs.

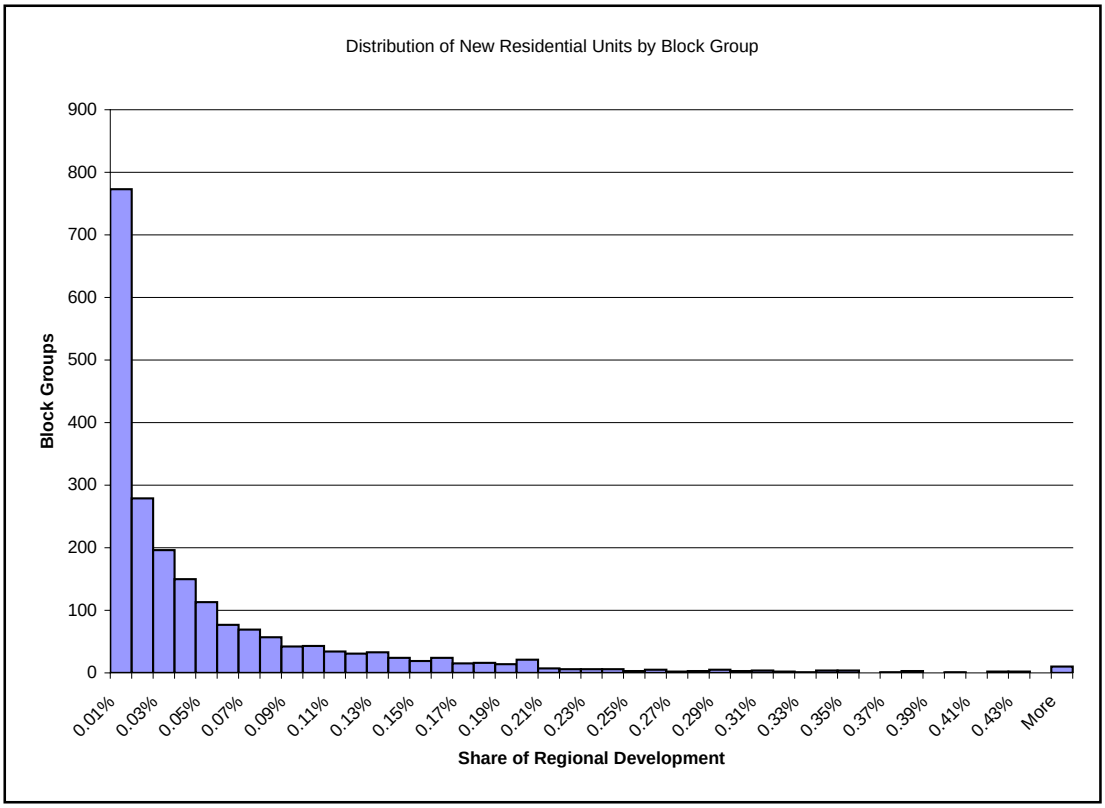


Figure 2.3. Block Group distributions of new residential housing units.

Rural, the first stage, is characterized by a low density of housing, low construction activity, and predominately single family units. Suburbanization stage would normally come next as the first wave of development occurs and the rural areas experience high rates of residential development activity. Infill, the next stage, is a slower pace of growth of land development with increased land values, housing unit values, and increased construction of multi-family structures.

The fourth stage, Packing, pushes population densities to their maximum resulting in overcrowding, the conversion of larger homes into multi-unit housing, and a diminution of new construction. Thinning, the fifth stage, identified a neighborhood characterized by declining populations and increased vacancies, fueled by

deteriorating housing conditions and abandonment. In the final stage, Recapture, housing units are either rebuilt or rehabilitated, increasing neighborhood population growth and densities.

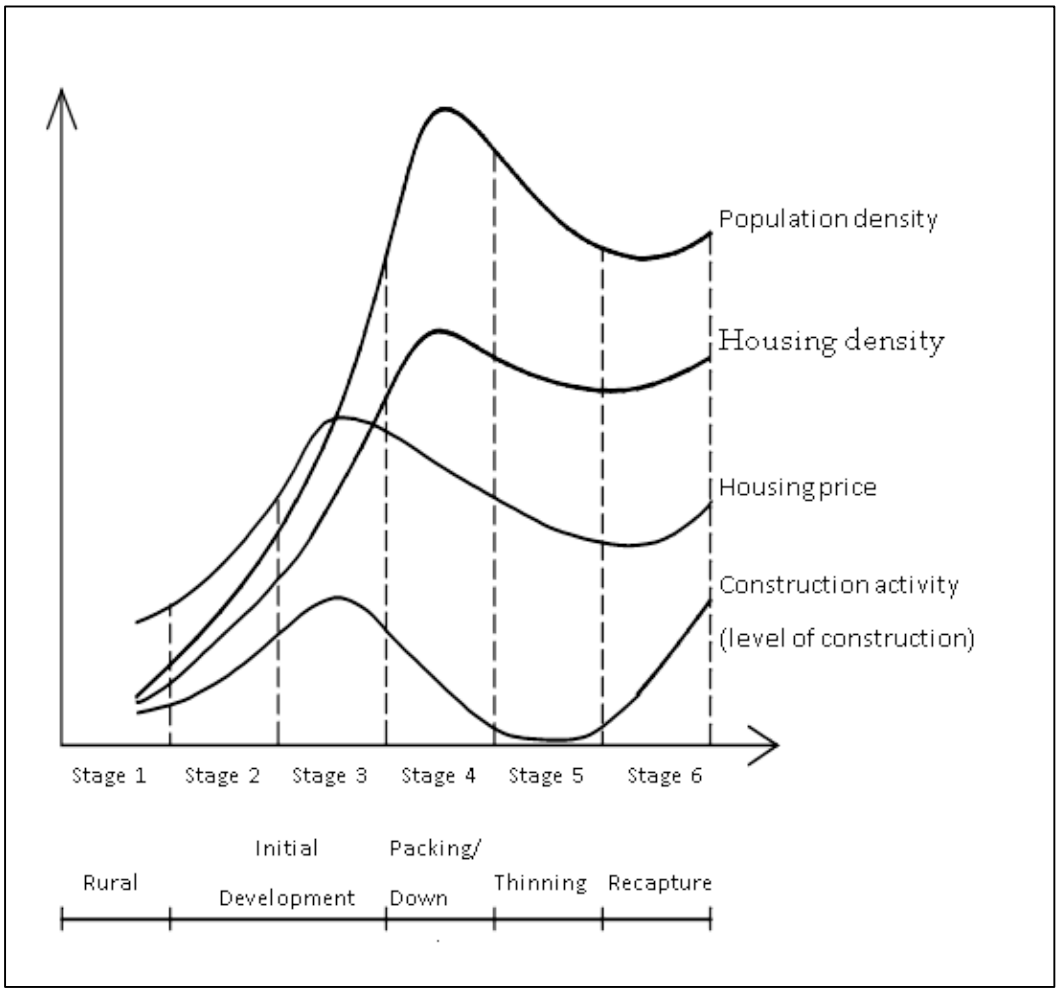


Figure 2.4. Birch's Neighborhood Housing Stages.

Characterizing neighborhood change by "stages" gives a false impression that this is an evolutionary continuum and that all neighborhoods pass through each stage. However, this is not the case. Skipping of stages often occurs. However, despite the

lack of strict linearity, use of these stages helps us in classifying BGs based on their current level of development, and probable transitions.

My concern for this study are areas previously rural or currently identified as low density. Therefore, the focus is on BGs within neighborhood stages two or three, Suburbanization or Infill respectively. Although stage six, Recapture, also includes new housing unit construction, it occurs on lands already classified as urban and unlikely providing desirable habitat.

Using Census1990 and Census 2000 data, the neighborhood stage of each BG was computed. The 1990 stage scores modeled the share of residential development in a Block Group expected from 1990 to 2000. The 2000 stage scores were modeled to predict BGs likely to receive substantial shares of the regions forecasted new housing from 2000 to 2010.

Independent variable: new housing units in prior decade

Development inertia is a way to identify the potential BG recipients of new housing. Development inertia is based on the concept that areas experiencing substantial development are likely to continue on their growth path. Conversely, areas with little or no residential development occurring, are likely to continue to experience little or no development.

Although the prior level of new housing construction is a variable used in the neighborhood stage calculation, its use as a separate variable gives the momentum of development greater weight when taking place in suburbanizing areas.

Independent variable: regional labor market area

The USDA has developed, based on commuting patterns, economic groupings of counties into labor market areas or LMAs (Tolbert & Killian, 1987; Tolbert & Sizer, 1996). The ten counties of my study area are included within four of the 394 LMAs of the United States as shown in Figure 2.5. For purposes of this study, I grouped Rockland County with LMA 194, the same LMA which includes the counties of Putnam and Westchester. An evaluation of population density and housing values for each Rockland County BG relative to the other BGs within the LMA 194 region provided the basis for this re-grouping. Rockland County shared similar levels of population density and similar housing values as Putnam and Westchester counties suggesting the grouping of Rockland County with LMA 194 did not bias the results. The independent variable of LMA designation was included to test for a regional effect on residential development. The relationship was small but significant.

Independent variable: density of local road network

The Road Density layer's purpose is to identify BGs with a strong existing transportation network to provide the access necessary for development. It is intuitive that land parcels with existing road accessibility have a lower development cost. To a limited extent, the road density layer is also a surrogate for local zoning. The linkage between road density and local zoning can be explained as follows: 1. Public roads must be officially accepted by the municipality, resulting in the municipality assuming ownership and maintenance responsibilities; 2. Municipal public works projects are designed to address existing problems or predicted needs and must meet existing

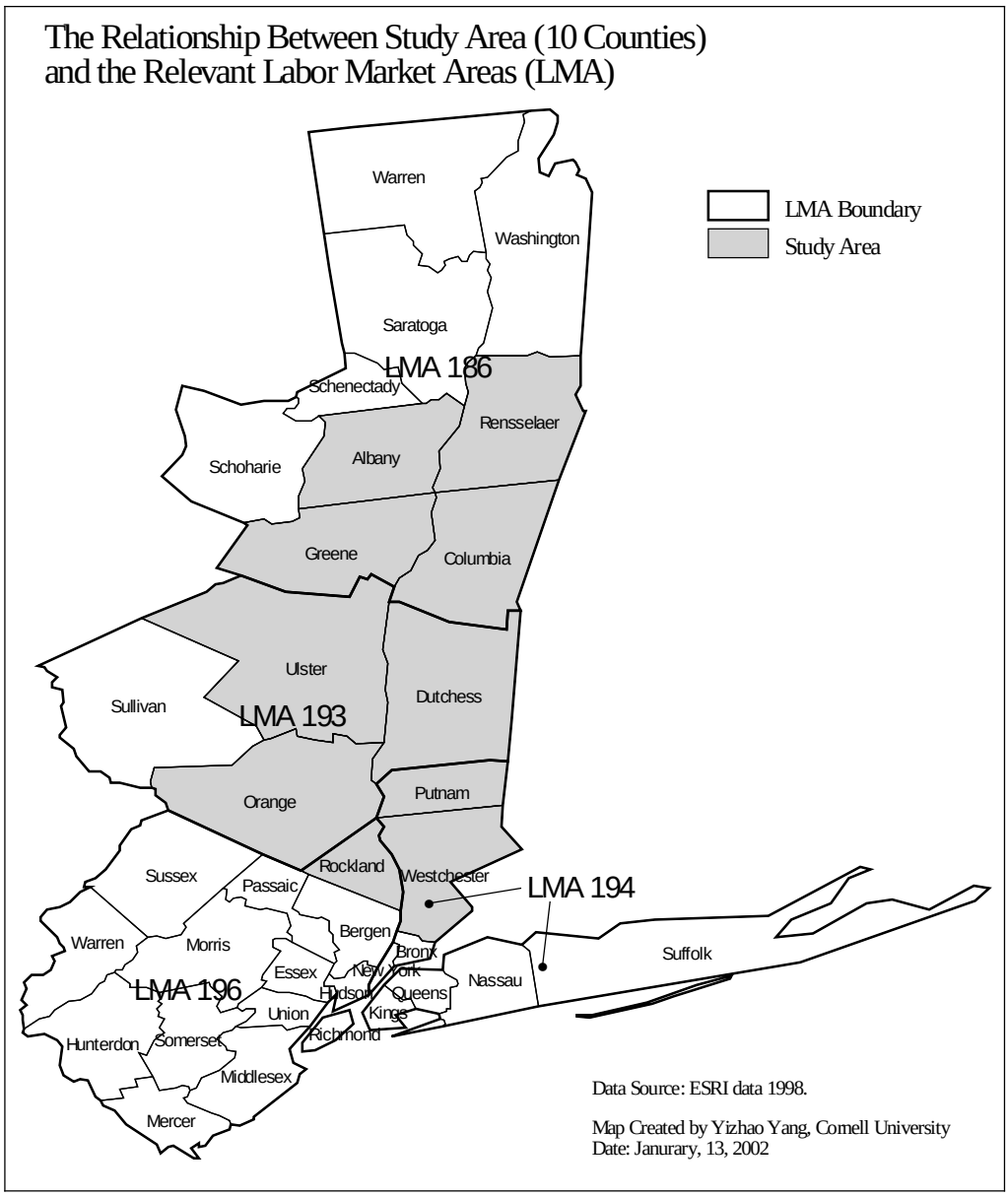


Figure 2.5. The relevant labor market areas in the study area.

zoning restrictions; 3. Public utilities, such as water and sewer, often utilize the right-of-way of public roads. Therefore, road density was deemed a reasonable surrogate for local land use zoning. This surrogate was necessary because the principle of

"home rule" is supported by New York State. Home rule vests the "lowest" or most local governmental body (town, city, or village) with zoning. The municipal fragmentation of the Hudson River Valley into 164 towns and cities creates the potential for 164 different sets of zoning classifications and regulations. The task of creating a single digital zoning database would be daunting and exceed both the financial and time constraints of this study.

The road density dataset, developed from the 1994 TIGER data, includes correction and updates to the 1990 release (USCB, 1994). Using the Census 2000 BG boundary delineations, road data along the boundaries were removed from the density calculations by using a 5-meter buffer on each side of the BG boundary and erasing all road segments within the buffer. This removal was done to acquire a measure of the BG's internal fragmentation by eliminating roads which were also boundaries. The cross-tabulation of road length and BGs was prepared to yield the sum of road length (m) per BG area (sq. m).

Independent variable: proximity to centers of population

To quantify the proximity of each BG to a select group of twenty regional centers of population, I used a measure called population potential (Edmonston, 1975). Population potential measures the proximity of a place to points of population concentrations. Population potential for each population center at a single BG is the ratio of the population of a selected center to the distance between the BG and that center. The theory is that the impact of a population center is directly proportional to

its size and proximity. The sum of all the ratios for all population centers is calculated to acquire a total population potential for the BG.

Investigation of physiographic data

The BG road density variable was the only physiographic factor to be included as an independent variable. Other physiographic features such as proximity to road intersections, slope, wetlands, open water, and stewardship status were investigated but failed to be significant. However, the identification of developable land within a BG utilized some of these factors, potentially affecting the receptivity of the BG to new housing.

The Census 2000 data includes land area values for each BG derived from the calculated area of a BG and the open water area as delineated by the TIGER/Line files (USCB, 1994). I generated a "developable land" area based on data from New York State Department of Environmental Conservation (NYSDEC) and NY-GAP project (C. Smith et al., 2001a).

Before determining a BG's developable land, it was necessary to identify areas with legal or physical constraints which would make development difficult. In this study, these areas include: open water areas mapped in either the Census 2000 data layer (USCB, 2002) or the NY-GAP data layer (C. Smith, 2001b); land areas with 30% or greater slope gradients (American Society of Planning Officials, 1959); areas designated freshwater wetlands by New York State (NYSDEC, 2013); and areas with development controls as identified by NY-GAP (C. Smith, 2001a). This development constraints layer was then intersected with the BG data layer to calculate the total

developable land for each BG in hectares. The generation of potential intensive residential growth areas (IGA) was based upon the calculated developable land area within the BG.

Determination of expected species biodiversity areas

I supplemented datasets of expected species distributions for amphibians, reptiles, mammals, and birds derived from the HRV-GAP data (C. Smith et al., 2001b, pp 53-64) with a dataset of the total aggregation of all vertebrate classes, developed from the same data source.

The 30x30 meter cell resolution data from NY-GAP contained an identifier for each unique combination of species and assigned that identifier to all grid cells with that combination. I compiled each vertebrate group dataset in this manner.

From the HRV-GAP expected species distribution data, a species richness value was calculated for each of the Block Groups in the study area and each vertebrate species class (i.e., amphibians, reptiles, mammals, and breeding birds). The specifics of the above process can be obtained from C. Smith et al., 2001b (pp. 46-48).

Results

Identification of potential intensive residential growth areas

The five variable binomial model explained 37 percent of the variation in the level of residential development and provided a probability of development values for each BG. I identified 77 of the 2,212 BGs in the ten counties of the HRV as prime candidates to receive major shares of new housing forecasted to be constructed

between 2000 and 2010. These are the IGAs for the highest levels of expected residential development (Figure 2.6). Another 585 BGs have conditions also making them likely to receive substantial portions of new housing units forecasted for the region. Collectively, these 662 BGs form the potential intensive growth areas (PIGAs). To associate these PIGAs with counts of new housing units, the experience of the previous decade, 1990 – 2000, is used. Over the previous decade, US Bureau of the Census reported the construction of 89,648 new housing units in the HRV (USCB, 2000). Based on the past, if the level of residential development remains the same in this decade, then it is likely that the IGA BGs may receive 135 or more new HUs and the 585 PIGA BGs may receive 90 to 135 new HUs each.

The IGAs are BGs with some factors favorable to residential development. These factors include relatively high levels of recent new housing construction, access to population centers, stage of neighborhood development, and an existing system of roads. I have identified the BGs within the region where these factors substantially favor residential development. Because a forecast of anticipated housing growth existed, I modeled shares of residential development likely to occur in each BG. Because of the existance of a regional forecast of expected housing growth, my share predictions can be used to distribute the expected regional HUs down to the BG level.

According to the Censuses of 1990 and 2000, the number of new housing units was 125,939 for the 1980s and 89,648 for the 1990s (S. Smith et al., 2004, p. 2). The predictions focus on the BGs within the HRV where conditions are favorable for sizable residential growth.

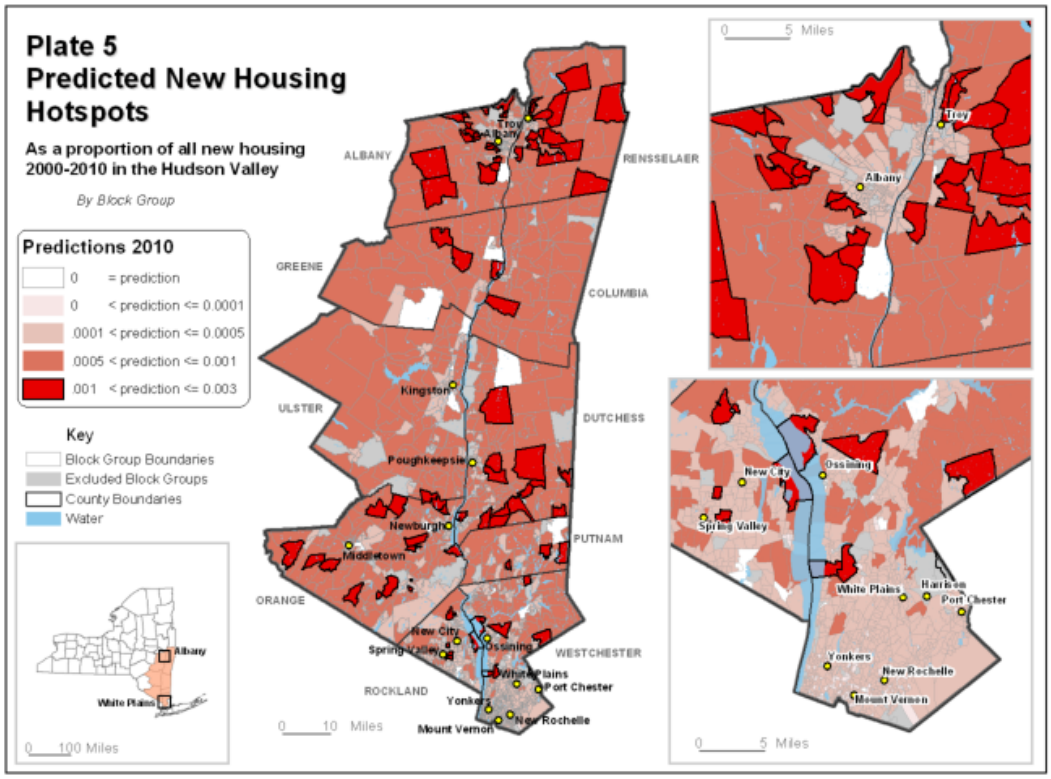


Plate 5. U.S. Census Block Groups in the Hudson Valley Study Area showing the greatest predicted new housing development by 2010. Predicted new housing development is expressed as a proportion of all predicted new housing in the Hudson Valley Study Area for each Block Group. Seventy-seven Block Groups have the greatest potential for new housing development by 2010 (outlined in bold).

Source: U.S. Department of Commerce U.S. Census Bureau Geography Division, Census Blocks, New York State (Shapefile: 2001); Smith, S.D., W.A. Brown, C.R. Smith, and M.E. Richmond, 2004. Habitat Vulnerability Assessment for the Hudson River Valley, Final Report, Hudson River Foundation, New York, NY.

Figure 2.6. Predicted new housing development for BGs in the study area. Values greater than .001 are IGAs (referred to as Hotspots in Plate 5). Values greater than .0005 and less than or equal to .001 are PIGAs. Extracted with permission from S. Smith et al., 2004, pp 17.

The BGs, within an IGA or PIGA, are distributed throughout the study area (Figure 2.7). The highest and lowest number of PIGA's occur within Orange and

Greene Counties, respectively. The highest and lowest number of IGA's occur within Albany and Ulster Counties, respectively.

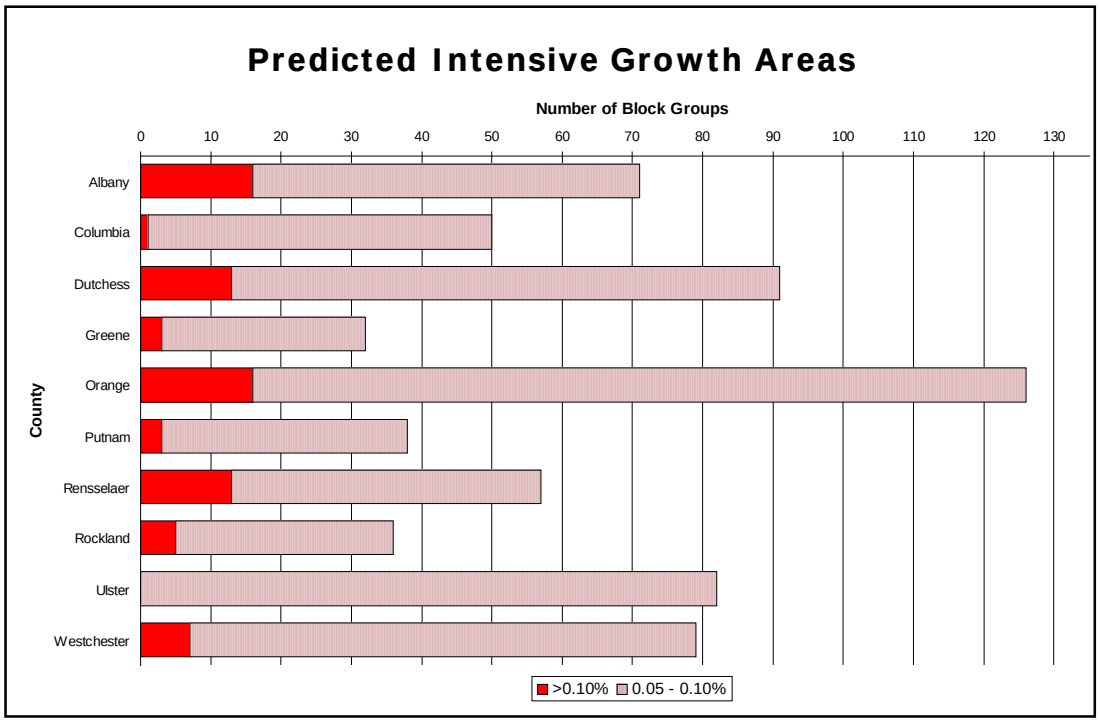


Figure 2.7. Distribution of predicted intensive growth areas and potential growth areas by BG for each county in the study area.

Identification of expected biodiversity richness areas

The ratio of all species expected within the BG and all species expected within the HRV determines the species richness value for each BG. These ratios permit comparison across vertebrate groups by providing a relative concentration rating for each species group. For more detailed species information, including species mapping by quartiles within BGs, see S. Smith et al., 2004 (pp 18-27). Table 2.1 shows the specific quartile breakpoints for each of the vertebrate groups.

A review of Table 2.1 gives a false impression that the total number of expected vertebrate species for the study area is approximately 68. The HRV-GAP reports the expected species totals for each vertebrate group as follows: Total Vertebrates = 308, Amphibians = 28, Reptiles = 27, Mammals = 54, and Breeding Birds = 199. The HRV-GAP reports total percentages for expected listed Threatened, Endangered, and Sensitive species as 80% of Total Vertebrates, 78% of Amphibians, 56% of Reptiles, 100% of Mammals, and 86% of Breeding Birds (C. Smith et al., 2001b, pp iii). The lower numbers presented in Table 1 are a function of the weighting process used in calculating the BG species number. For a detailed explanation of the weighting process, see C. Smith et al., 2001b, Appendix.

Table 2.1 indicates that amphibians, mammals, and breeding birds have significant numbers of BGs with low concentrations as evident by the wide range of species counts in the first quartile. Reptiles appear to be more evenly distributed within the first quartile BGs. The Threatened, Endangered, and Sensitive vertebrate species groups repeat the same pattern.

Table 2.1 also reveals an interesting effect for the breeding birds species which also mimic the weighting effect referenced earlier when discussing the total vertebrate species count. For the breeding birds vertebrate group, the largest vertebrate group in the study area, 199 species are expected (C. Smith et al., 2001b), and yet the maximum range shown in the 4th quartile is 37.6. This low species concentration number is a function of the weighting scheme discussed in the Methodology section. The spatial concentration of bird species within their habitats aggravates this effect. This concentration effect would result in many instances of adjacent 30x30 meter grid

cells being occupied by different species but receiving the same species count, resulting in the underestimation of the total species present. Because relative concentrations are what interests us, this underestimation is not a major concern.

Table 2.1. Species richness quartile breakpoints for each vertebrate group. Quartiles enable the comparisons of species concentrations across vertebrate groups.

Species Distribution Breakpoints

		Species Richness Quartile Breaks							
		1st Quartile		2nd Quartile		3rd Quartile		4th Quartile	
Plate Title	Plate #	Low	High	Low	High	Low	High	Low	High
Expected Vertebrates	6	5.6 --	34.8	34.9 --	36.5	36.6 --	38.8	38.9 --	43.9
Expected Amphibians	7	0.5 --	40.7	40.8 --	42.5	42.6 --	56.9	57.0 --	67.3
Expected Reptiles	8	14.3 --	32.2	32.3 --	33.1	33.2 --	42.5	42.6 --	51.4
Expected Mammals	9	0.2 --	47.5	47.6 --	50.0	50.1 --	57.2	57.3 --	67.9
Expected Breeding Birds	10	2.4 --	27.4	27.5 --	30.7	30.8 --	32.3	32.4 --	37.6
Expected TES Vertebrates	11	1.7 --	4.6	4.7 --	6.6	6.7 --	12.9	13.0 --	17.0
Expected TES Amphibians	12	0.6 --	13.1	13.2 --	14.2	14.3 --	14.3	14.4 --	27.9
Expected TES Reptiles	13	0.0 --	10.0	10.1 --	19.4	19.5 --	29.4	29.5 --	52.6
Expected TES Mammals	14	0.0 --	33.1	33.2 --	33.3	33.4 --	46.1	46.2 --	65.3
Expected TES Breeding Birds	15	0.0 --	0.9	1.0 --	2.7	2.8 --	7.3	7.4 --	12.8

Intersection of Intensive Growth Areas and Expected Biodiversity Richness Areas

The relative concentrations of total expected vertebrate species across the Hudson River Valley, shown in Figure 2.8, are overlain with the 77 IGA BGs.

The expected pattern of lower species concentrations near the metropolitan areas is evident in (Figure 2.8). Many BGs with shoreline and associated with population centers are also ranked low. Some large BGs in Greene, Ulster, and Orange counties, associated with the Catskills Park Preserve and state parks, do not follow this pattern, but exhibit lower total vertebrate rankings than expected.

Species distributions by quartiles within the 77 BG identified intensive growth areas for each vertebrate group, and each TES vertebrate group, show that 53% of the

vertebrate species are within the IGAs (Table 2.2). It is also evident that 41% of the expected TES vertebrate species are within the IGAs.

Table 2.2. Species richness distribution within the 77 Block Groups identified as intensive growth areas. Quartiles enable the comparisons of species concentrations across vertebrate groups.

Species Distribution Within 77 Residential Development Hot Spot Block Groups

		Species Richness Quartile Based on All Block Groups in Study Area							
		1st Quartile		2nd Quartile		3rd Quartile		4th Quartile	
Plate Title	Plate #	Count	%	Count	%	Count	%	Count	%
Expected Vertebrates	6	6	7.8	10	13.0	20	26.0	41	53.2
Expected Amphibians	7	3	3.9	4	5.2	33	42.9	37	48.1
Expected Reptiles	8	7	9.1	2	2.6	30	39.0	38	49.4
Expected Mammals	9	3	3.9	8	10.4	19	24.7	47	61.0
Expected Breeding Birds	10	19	24.7	44	57.1	9	11.7	5	6.5
Expected TES Vertebrates	11	2	2.6	3	3.9	40	51.9	32	41.6
Expected TES Amphibians	12	49	63.6	11	14.3	1	1.3	16	20.8
Expected TES Reptiles	13	0	0.0	8	10.4	32	41.6	37	48.1
Expected TES Mammals	14	3	3.9	2	2.6	22	28.6	50	64.9
Expected TES Breeding Birds	15	2	2.6	2	2.6	39	50.6	34	44.2

Figure 2.9 shows the overlay of the 77 IGA BGs onto the total expected vertebrate distributions for Threatened, Endangered, and Sensitive (TES) species in the study area. The pattern of fewer species near metropolitan areas is evident. However, there is also a pattern of higher total TES vertebrate species totals in many BGs along the river starting at the southern part of the valley, tapering to fewer Block Groups as it travels northward. Most of the HRV study area BG concentrations of TES species fall in the third and fourth quartiles implying that the HRV contains more than 50% of the total TES vertebrate species expected statewide. The other BG quartiles are heavily concentrated near the urban areas and comprise a smaller total area. More detailed TES species information by BG can be obtained in Smith et al. 2004, pp 18-37.

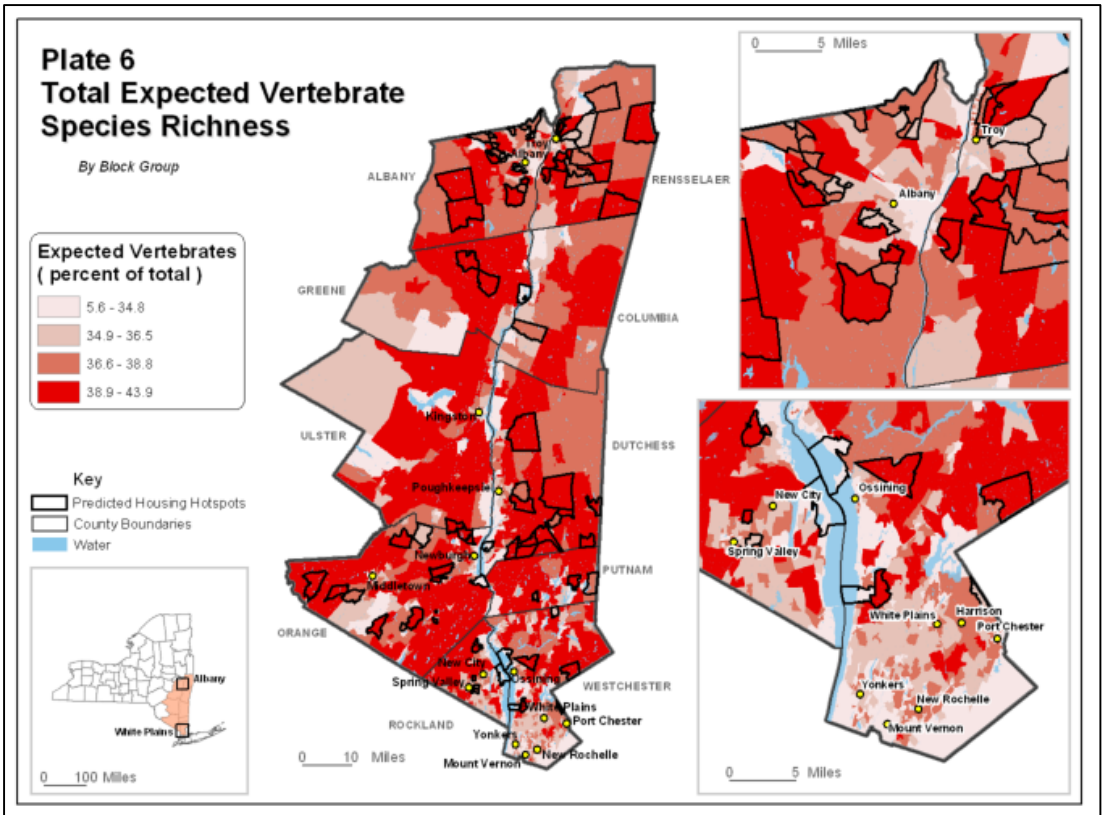


Plate 6. Total expected terrestrial vertebrate species richness, based upon wildlife habitat association models (see Appendix A for explanation) for Block Groups in the Hudson Valley Study Area. Because Block Groups are of different sizes, expected species richness is expressed as a relative density, where Relative Density = (Count of Expected Species for a Block Group / Block Group Area) x Total Number of Species Expected in the Hudson Valley Study Area. Areas of greatest potential conflict are represented by those Block Groups with the greatest potential for new housing by 2010, outlined in black (see also Plate 5) on the map. Areas of potential conflict merit more detailed study to verify species occurrences within their boundaries.

Figure 2.8. Spatial distribution of the IGAs by BG overlaid onto total expected vertebrate species richness by BG. Extracted with permission from S. Smith et al., 2004, pp 19.

Discussion

When evaluating the intersections of the areas of residential growth and species richness, it is best to bear in mind that these areas are the result of two separate predictions, each with their assumptions and possible errors. The first prediction is of

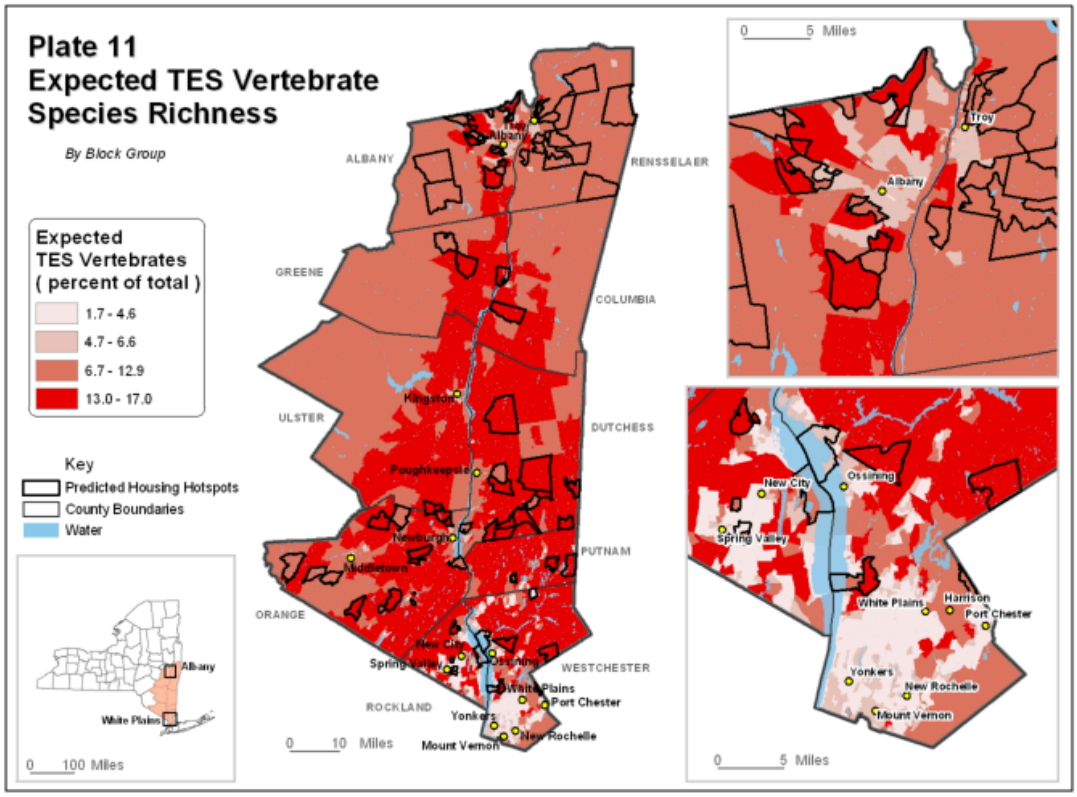


Plate 11. Total expected richness for TES (New York State listed endangered, threatened, and special concern) species of terrestrial vertebrates, based upon wildlife habitat association models (see Appendix A for explanation) for Block Groups in the Hudson Valley Study Area. Because Block Groups are of different sizes, expected species richness is expressed as a relative density, where Relative Density = ((Count of Expected Species for a Block Group / Block Group Area) / Total Number of Species Expected in the Hudson Valley Study Area) * 100. Areas of greatest potential conflict are represented by those Block Groups with the greatest predicted new housing development by 2010, outlined in black (see also Plate 5) on the map. Areas of potential conflict merit more detailed study to verify species occurrences within their boundaries.

Figure 2.9. Spatial distribution of the IGAs by BG overlaid onto total expected TES vertebrate species richness by BG. Extracted with permission from S. Smith et al., 2004, pp 29.

housing growth based on past trends and the second is a prediction of biodiversity patterns derived from expected occurrences of terrestrial vertebrates by a spatially explicit vegetation associations model (S. Smith et al. 2004, pp 18-37).

Even with the above caveats, it is evident that these results support the original hypothesis at both levels: 1) The spatial distribution of residential housing growth can be modeled at the Census BG level with Census data and other publicly available data, and 2) The results can provide a valuable coarse filter for the identification of biodiversity areas at risk from projected residential growth.

Local planners choosing to take actions to reduce potential conflicts between housing growth and biodiversity conservation would be well served to perform field studies to confirm occurrences of predicted species within their towns or counties based on the IGA's and PIGA's identified (S. Smith et al., 2005).

An examination of the vertebrate group maps provided in the Habitat Vulnerability Assessment for the Hudson River Valley report reveals that the distribution highlights the intra-group pattern (S. Smith et al., 2004, pp 20-27). Also, there is similar variation among the TES species groups (S. Smith et al., 2004, pp 30-37). Species with the TES designation currently enjoy special legal protection under the NYS Environmental Conservation Law. This protective status increases the desirability to avoid or mitigate potential conflicts with housing development.

All of the data required by the model were acquired or derived from readily available public data at a spatial resolution appropriate for regional studies. Development of the Birch's Neighborhood Stages required the collection of detailed housing permits and similar local non-digital data. Although this did not meet the goal of all data being available in digital format, it was required due to the timing of the study. These data are readily available at the county level, helping to retain the model's applicability to other regions in New York or other states. As a means to

identify areas of potential concern, the model is a practical and useful tool for county and regional planning.

References

American Society of Planning Officials. (1959). *Hillside Development. Planning Advisory Service. Information Rpt.* (No. 126). Chicago, IL: Clark, R.A.

Birch, D.L. (1971). Toward a Stage Theory of Urban Growth. *Journal of American Institute of Planners*, 37(2), 78-87.

Edmonston, B. (1975). *Population Distribution in American Cities*. Lexington, CT: Lexington Books.

Hoover E. M. & Vernon R. (1959). *Anatomy of a Metropolis: the changing distribution of people and jobs in the New York Metropolitan Region*. Cambridge, MA: Harvard University Press.

NYSDEC (2013). *New York State Regulatory Freshwater Wetlands, 1999-2013*. Albany, NY: NYS Dept. of Environmental Conservation. http://cugir.mannlib.cornell.edu/READMES/fwrdme.html

Scott, J.M., Davis F., Csuti B., Noss R., Butterfield B., Groves C.,…Wright R.G. (1993). *Gap Analysis: A Geographic Approach to Protection of Biological Diversity*. Wildlife Monographs 123:1-41.

Smith, C.R., DeGloria S.D., Richmond M.E., Gregory S.K., Laba M., Smith S.D.,…Weber J.T. (2001a). *A Gap Analysis of New York NY-GAP. Final Report and Data*. U.S. Geological Survey, Biological Resources Division, National Gap Analysis Office, Moscow, ID. 2-Disc CD set (issued October 2001).

Smith, C.R., DeGloria S.D., Richmond M.E., Gregory S.K., Laba M., Smith S.D.,… Hill E.A. (2001b). *An Application of Gap Analysis Procedures to Facilitate Planning for Biodiversity Conservation in the Hudson River Valley, Final Report, Part 1: Gap Analysis of the Hudson River Valley and Part 2: Atlas of Predicted Ranges for Terrestrial Vertebrates in the Hudson River Valley*. New York Cooperative Fish and Wildlife Research Unit. Ithaca, NY:Department of Natural Resources, Cornell University.

Smith S.D., Brown W.A., Smith C.R., & Richmond M.E. (2004). *Habitat Vulnerability Assessment for the Hudson River Valley: Final Report.* New York Cooperative Fish and Wildlife Research Unit. Ithaca, NY:Department of Natural Resources, Cornell University.

Smith S.D., Brown W.A., Smith C.R., & Richmond M.E. (2005). *Habitat Vulnerability Assessment in the Hudson River Valley*, GAP Analysis Bulletin No. 13, December, 37-43.

Tolbert, C. M. & Killian M. S. (1987). *Labor Market Areas for the United States*. Staff Report No. AGES870721. Washington, DC: Economic Research Service, USDA.

Tolbert C. M. & Sizer M. (1996). *U.S. Commuting Zones and Labor Market Areas: a 1990 Update*. U.S. Department of Agriculture Economic Research Service Staff Paper No. 9614. September, Washington, DC: Economic Research Service, USDA.

USCB (1994). Geography Division New York State. TIGER/Line Files. Updates to 1990, U.S. Census Bureau, Washington, DC.

USCB (1999). Housing Unit Review File. New York State. Population Estimates Branch, U.S. Census Bureau, Washington, DC.

USCB (2000). Census 2000 Summary File 3A, New York, U.S. Census Bureau, Washington, DC.

USCB (2002). Census 2000 Summary File 3. Technical Documentation. U.S. Census Bureau, Washington, DC.

Yang, Y., Brown W., & Smith S. (2001). *Using Neighborhood Growth Stage Model to Analyze New Residential Development in Ten Counties in Hudson River Valley Area, New York State.* Population & Development Program Working Papers. Ithaca, NY: Program in Applied Demographics, Cornell University.

CHAPTER 3:

MODELING RESIDENTIAL GROWTH AND HABITAT VULNERABILITY ON A REGIONAL SCALE

Abstract

A broadly applicable, geospatial residential growth model based on Block Group (BG) Census data has been developed to identify potential intensive residential growth areas (IGA). A linear regression model was used to explain residential housing unit increases over two ten-year periods. The model used five BG-based variables: current housing unit density, current road density, previous decade's housing unit change, housing unit density gradient derived from adjacent BGs and population potential index. The model, drawn from a ten county region in Eastern NY extending from New York City to Albany, explained 38% of the variance associated with the identification of intensive residential growth areas and identified the top 5% of block groups showing increases in residential housing units over the last decade. Of the BGs predicted to be areas of fast growth, 53% and 41% were IGAs as computed from 2000 and 2010 Census data, respectively. The results of this broadly applicable regional model were applied towards the goal of biodiversity conservation as measured by species richness. Aggregated vertebrate species richness data facilitated the identification of species richness areas when overlaid with block group boundaries, resulting in the identification of 211 species-rich BGs, based on total vertebrate species counts. Of the IGAs predicted for 2000 and 2010, 16% and 8%, respectively, were also species-rich Block Groups. Early identification of biodiversity areas that are

at risk from potential residential growth can empower decision makers to address proactive education and planning, as well as critical mitigation options.

Introduction

The adverse effect of sprawling urbanization on ecologically sensitive areas has prompted a growing effort to understand spatial patterns of residential development (Burchell & Mukherji, 2003). My goal is the development of a census-based model that empowers local officials with vital information on potential areas of future residential growth and to highlight locations where such development will affect vulnerable habitats and animal species. I hypothesize that residential growth occurs in patterns capable of being modeled at regional scales using variables derived from Census and existing infrastructure data. A census-based model allows one to consider both social and economic factors that can affect residential growth using readily accessible data ensuring portability to any region of the United States. The National GAP Analysis Program offers the necessary inventory of rare, endangered and threatened species across all states to facilitate portability nationwide.

Identification of potential growth areas at a regional scale allows resource managers, conservation groups, and government agencies to focus their educational and regulatory efforts to preserve habitats and species diversity. This ability is critical in regions of species richness, such as the Hudson River Valley of New York State. Located between two of the State's largest residential housing generators, New York City and the City of Albany, this valley region provides suitable habitat for more than three-quarters of the terrestrial vertebrate species found within New York State (C.

Smith et al., 2001b). The ability to identify areas that are directly or indirectly impacted by residential growth and ascertain their association with species richness areas provides the capability to rank species richness areas at risk, focus efforts and thereby increase the effectiveness of associated conservation efforts and expenditures.

Direct impacts from residential development, such as land use conversion and habitat loss, will be more readily identified and, perhaps, less disruptive since mitigation measures can be readily assigned.

Estimates of residential development growth in the United States for 2000-2025 place a total land conversion to residential at 18.8 million acres at an average rate of 0.6 acres per residential unit (Burchell & Mukherji, 2003). However, indirect impacts (e.g., increased demands for infrastructure and public facilities such as roads, water, sewer, schools and municipal buildings), which are not necessarily spatially coincident with the land use conversion and habitat loss, are likely to be less obvious and could potentially be more disruptive. For example, increased residential development may over-burden an existing school district, leading to the need for additional capital expenditures on infrastructure (e.g., buses, buildings, and utilities) and causing district-wide impacts such as increased traffic and tax burden. The potential indirect costs of the estimated residential growth in the United States for the 2000-2025 period were also estimated (based on public water and sewer, roads and public service) at $1,117 billion dollars or $47,530 per residential unit (Burchell & Mukherji, 2003). Estimates of the potential acreage of additional land conversion are unavailable. The ability to anticipate where residential housing is likely to occur will also enable officials, managers, and stakeholders to undertake conservation efforts such as species

monitoring programs, education efforts and mitigation strategies related to direct and indirect impacts. Targeting these activities to smaller regions will increase the applicability and efficacy of the conservation efforts (e.g., monitoring activities can be confined to a select region and limited to known or predicted species within the area, thereby resulting in a reduction of labor-intensive activities, monetary investment and, likely, less controversy). Spatially- or taxonomically- extensive monitoring efforts are very time consuming and expensive (Perkins et al., 2013). Providing this cost-effective method of targeting conservation and monitoring efforts, may be the only way these programs can be implemented.

In an attempt to understand the dynamic spatial patterns associated with sprawling urbanization, efforts have focused on simulating future expansion of urban areas or build-out scenarios for counties and similarly sized areas (Birch, 1971; Ottensmann, 1977; Hawbaker et al., 2005).

Many Census-based studies have focused on population growth using build-out scenarios drawing from comprehensive plans. Developing these build-out scenarios requires the substantial input of localized data. This local data requirement is necessary for states such as New York where “home rule” places land use control in the hands of local government entities such as towns, cities, and villages (NYDOS, 2001). This situation dramatically increases the effort required to apply a build-out scenario methodology to a regional area. For example, the Hudson River Valley region includes 145 towns, 18 cities, and 95 village entities, some of which do not have zoning or comprehensive plans. To compound the task, zoning and comprehensive development plans are notoriously unique to each municipal entity.

Another issue associated with building projections upon these land use control mechanisms is that they are subject to political influences and circumvention. I sought to find a reasonable method of predicting residential housing growth using inputs that may be influenced by zoning and comprehensive plans, but are themselves independent.

Coupled with the need to predict growth areas is the identification of species richness zones. There have been species risk studies focused on the identification of habitat risk areas (e.g., White et al., 1997) which have utilized county-based and municipal-based development plans to anticipate growth patterns. Others have created systems to provide species data to inform planning decisions (e.g., Theobald et al., 2000). These efforts have been successful when accepted and fully implemented. The first type required substantial effort to create countywide development scenarios based on existing comprehensive plans and, as such, is limited in scope and not easily portable to other locations. The second type offered internet access to species distribution data to support the planning process. This process is much broader in scope and transferable, but again hinges on dedication to the planning process and continued follow-up about the developed plan. In a region where large areas are potentially ecologically important, and there is substantial development pressure, the political will to implement spatially extensive restrictions may not exist.

My goal was to develop a census-based model that empowers local officials with vital information on potential residential development locales and to highlight where such development will affect vulnerable habitats and animal species. A census-based model allows one to consider both social and economic factors that can affect

residential growth using readily available data (e.g., Census data for model variables), thus ensuring portability to any region of the United States.

Based on my hypothesis that residential growth occurs in patterns capable of being modeled at regional scales using variables derived from Census and existing infrastructure data, I identified three objectives. The first objective was to develop a census-based model to predict areas of rapid residential housing growth during the next decade within the BG geography. The second objective was to identify the spatial extent of regions with high species diversity and conservation significance. The third objective was to determine the spatial overlap of predicted residential growth areas with areas of species diversity, and provide a ranking of these zones. This novel approach easily integrates both biological and human demographic elements into the process of identification of natural areas at risk from residential development.

In this study, I develop an easily implemented, transportable method for modeling these growth patterns that enables the prioritization of at-risk biodiversity areas as defined by the total of all predicted species. The ability to predict the spatial distribution of residential development and identify ecologically sensitive areas susceptible to developmental impacts can empower decision makers. This empowerment may lead to prevention or mitigation of future negative impacts, offering a way to maintain the desirable natural habitats and biological diversity of the region.

Study area

The study area is a ten-county region (Albany, Columbia, Dutchess, Greene, Orange, Putnam, Rensselaer, Rockland, Ulster, and Westchester) in the Hudson River Valley of New York State (Figure 3.1). Contained within a rectangle with corner coordinates of 74^0 50' W, 40^0 50' N (lower left) and 73^0 10' W, 43^0 0' N (upper right),

The study area extends from Albany County to New York City. This region includes approximately half of New York State's human population and provides habitat for hundreds of migratory and resident species of wildlife (C. Smith et al., 2001a). This region is facing a period of re-industrialization, residential development, and constant development pressure emanating from New York City northward and, to a lesser degree, southward from the city of Albany (Hu, 2000).

The most recent data available, developed by a regional gap analysis project shows that more than 80% of the terrestrial vertebrate species that enhance New York State with their presence, inhabit this ten county region. The valley provides habitat for 69% of all amphibian species (25 species) and 58% of all reptile species (28 species) found in New York (C. Smith et al., 2001b). Habitat for 87% of all breeding bird species (214 species) and 90% of all mammal species (57 species) is also provided (C. Smith et al., 2001b). Among terrestrial vertebrates, 75% have all or a significant portion of their range within the Hudson River Valley study area (C. Smith et al., 2001b).

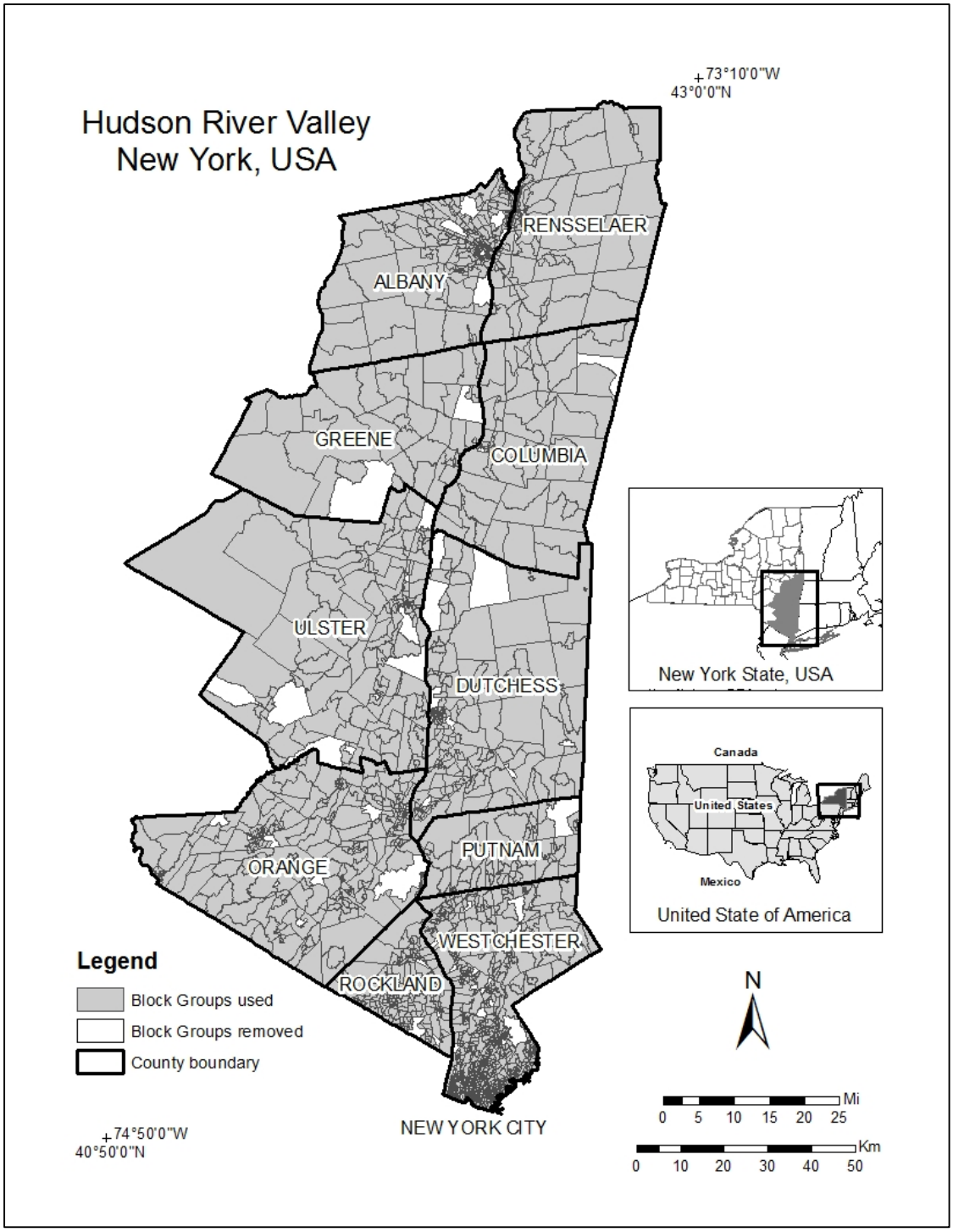


Figure 3.1 Ten-county region with the 2,111 Block Groups used in the analysis and the 101 omitted Block Groups that represented group quarters and institutional housing. Group quarters and institutional housing increases are subject to institutional and governmental forces outside the scope of this study.

Methods

Census BGs provided the geographical unit of analysis for the prediction of intensive residential development. Unlike political units such as counties and towns, BGs are statistical units of census geography. The boundaries of a BG identify an aggregation of the smallest Census geography, Blocks. Data from the U.S. Census Long Form questionnaires administered to a sample of households are tabulated for BGs and summarized by the U.S. Census Bureau (USCB, 1999; USCB, 2002). These areas contain approximately 400 housing units (range of 250-550 housing units). BGs may vary in extent and boundary delineation from one census to another (USCB, 1994). Because BG boundaries are variable from census to census, datasets of 1990 and 2010 Census data needed to be adjusted to the 2000 BG boundaries to provide comparable data, using Census Bureau derived methodology. Appropriately adjusted datasets, created by GeoLytics, Inc. (GeoLytics, 2000a; GeoLytics, 2000b), were obtained.

Initial model development involved 2011 of the 2212 Census BGs in the study area. A total of 101 Block Groups containing group-quarters, dormitories, and institutional housing were removed before model development because housing unit growth within these BGs is responding to very different factors than those of non-institutional residential housing. The random subdividing of the 2011 BG dataset provided the two datasets, comprised of 1058 and 1053 BGs respectively, used for model development and validation. The model's dependent variable is based on the total housing unit count of a BG.

Housing unit counts are more desirable than population counts as a representation of residential growth because they better represent the spatial impact on the environment (e.g., the footprint of the housing structure, transportation and utilities, and the social/commercial infrastructure) than population counts (Theobald, 2001; Hammer et al., 2004). According to the Census 2000 report, the ten counties of the Hudson River Valley contained 2,212 BGs with a size range of 0.02 to 544.16 km^2 and an average size of 7.37 km^2. Census 2000 also reports that the number of housing units per BG in the Hudson River Valley varies from a minimum of 0 to a maximum of 2,602, with a mean of 396 (USCB, 2002).

Census 2000 data on the year of housing unit construction determined the level of new residential development of each BG. Housing units classified as "new" had to have been built from 1990 to 2000. For areas the size of the study area's BGs, the ten-year interval of the decennial census data is advantageous for two reasons. First, the data were measured consistently and uniformly across housing units types and political jurisdictions. Second, the ten-year interval is longer than a typical business cycle and therefore the net change over the decade - new construction, conversion of existing units, and demolitions - are less representative of short-term fluctuations that might overemphasize a boom or bust period.

The skewed histogram of the raw housing unit counts for each BG (Figure 3.2a) necessitated the investigation of various transformations to identify a derivable normally distributed dependent variable. The final transformation used for the dependent variable of housing units was the natural log of the housing units for each BG (Figure 3.2b).

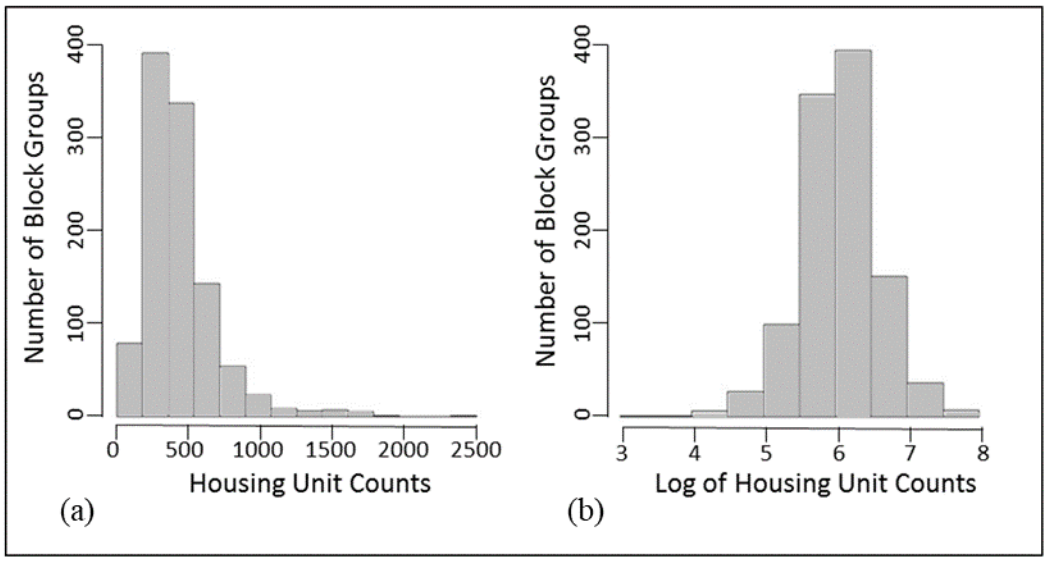


Figure 3.2 Histograms of: a) dependent variable and b) natural log transformation of the dependent variable.

Development of an earlier model (S. Smith et al., 2004) suggested the use of four independent variables. These four variables are housing unit density for previous decade (huden1990), road density for preceding decade (rdden1990), housing unit change occurring between the two preceding decades (huchng8090), and population potential for the preceding decade (poppo1990). Initial model development included the four previously identified variables in conjunction with six other new variables. Two of the new variables attempted to address localized influences, the difference between a BG's housing unit density and the mean housing unit density for all adjacent BGs (meandifhuden1990) and the difference between the BG's road density and the mean road density for all adjacent BGs. Two other new variables involved landscape factors impacting development potential, specifically, the percentage of a BG's area comprised of slopes more than 30% and the developable proportion of a

BG's area (portions not within wetlands, open water, public ownership, and excessive slopes). Protective status, one of the remaining two new variables, is based on the protective status of a BG's area as identified by the NY-GAP data. Lands already protected are classified in the stewardship class of 1, 2, or 3 (C. Smith et al., 2001a). Finally, the remaining new variable, the ratio of single-family housing units to multifamily housing units, is an attempt to address under-utilized housing units.

Comparing AIC values consecutively on each of the variations of a ten variable model (the original four plus the above mentioned six) resulted in an optimal model containing only five statistically significant predictor variables. Of the additional six variables investigated, only the meandifhuden1990 (the difference between a BG's housing unit density and the mean housing unit density for all adjacent BGs) showed any significance as a predictor of future housing unit development, resulting in its inclusion in the optimal model.

Predictor variables from the previous census, Census 1990, were used to explain variations in the amount of residential development among BGs. The model which best explained residential development variation had five predictor variables: housing unit change in the prior decade (huchng), road network density (roadden), housing unit density (huden), population potential (poppot), and a localized housing unit gradient (meandifhuden). Shown below is the linear equation used to predict the dependent variable loghu for 2000. An explanation of the predictor variables follows.

loghu2000 ~

huden1990 + roadden1990 + huchng8090 + poppot1990 meandifhuden1990

To help clarify, this equation can be presented as follows: log of housing units in 2000 = housing unit density in 1990 + road density in 1990 + housing unit change from 1980-1990 + population potential in 1990 + housing unit density difference from the adjacent BGs in 1990.

Predictor variable: housing unit density

Housing unit density (huden1990) is a single variable representing initial growth and is intended to mimic Birch's Neighborhood Stages (Birch, 1971). Housing unit density is a function of the total number of units reported in the current Census and the BG's area.

huden1990 = housing unit count / BG area

To provide a more accurate housing unit density, only the land area of the BG was used as calculated from the total area minus any recorded water area. No attempt was made to confirm or augment the recorded water area of any BG. As the housing density value increases, the BG moves away from a low density, perhaps rural neighborhood, towards a fuller-development stage.

Predictor variable: road network density

Road network density (roadden1990) is the total length of roads, at the time of the current Census, divided by the land area of the BG.

roadden1990 = road length / BG area

The purpose of the road network density variable is to identify areas possessing a substantial transportation network providing the necessary access for development.

Road density is positively correlated with housing density (Sanchez, 2000; Hawbaker et al., 2005). This positive correlation seems intuitive because land area which can be accessed by an existing road has a lower development cost and thereby is more likely to be developed. What is not as intuitive is at what point does higher road density inhibit development (i.e. Is there a level at which the fragmentation of the region negatively impacts future development)?

Predictor variable: housing units built in previous decade

This variable (huchng8090) measures the development inertia of a BG based on preceding decade's growth and is expected to be a positive impact during the subsequent decade's housing unit growth.

huchng8090 =

(current census housing unit count – previous decade's census housing unit count)

Simply stated, those BGs with little or no residential housing change in the preceding decade (1980 – 1990) are likely to continue to have little or no development. On the other hand, BGs that experienced high levels of residential development in the prior decade are likely to continue to be areas attractive to development.

Predictor variable: population potential

Population potential is an inverse distance weighted summation of the influence exerted by significant population centers upon each BG.

$$\text{poppot1990} = \sum_{j=1}^{n} \left(\frac{P1}{d1} + \frac{P2}{d2} + \frac{P3}{d3} \ldots + \frac{Pn}{dn}\right)$$

This variable, based on the current census, is the summation of the ratios of the population at all Census places to the distances from each Census place to the specific BG. This variable is expected to be negatively correlated with the independent variable (Gustafson et al., 2005).

Predictor variable: local housing unit density gradient

This variable (meandifhuden1990) was derived from the current Census data and is intended to provide a spatial measurement of localized development pressure.

$$meandifhud1990 =$$

$$\sum_{n=1}^{n} (hudenadjBG1 + hudenadjBG2 \ldots + hudenadjBGn)) * 1/n\ )$$

$$- huden$$

This variable is the summation of the housing unit densities, based on the current census, for each BG adjacent to a specific BG.

The premise is that the movement of neighboring BGs towards residential development will tend to pull others along (Theobald & Hobbs, 2002). Conversely, their reluctance to develop will tend to constrain others from developing.

Determination of species risk areas

Maintenance of the terrestrial vertebrate species richness, as defined by the species count associated with a given BG, was a goal. Achieving the maintenance of the terrestrial

vertebrate species richness requires the identification of BGs likely to receive an increase in housing units and their relationship to species richness determined. At that point, local planners desiring to incorporate conservation of biodiversity into the planning process can mitigate impacts by directing development activities away from potential development BGs that exhibit higher species richness.

Development of total expected vertebrate class distributions data sets used expected species distribution data from the HRV-GAP project (C. Smith et al., 2001b). The HRV-GAP project provided raster (grid cell) data, at a 30x30 meter cell resolution, for amphibians, reptiles, mammals, resident birds, and an aggregation of all four vertebrate. Expected species were those likely to be present as determined by the species habitat association models developed for NY-GAP (C. Smith et al., 2001a). Block Groups with total species counts placing them in the top 10% of all BGs were identified as species concentration areas.

To assign a species count number to an entire BG required the extraction of all grid cells with 50% or more of their area within the BG. Each unique species count for the grid cells extracted was weighted by the total cell area of the BG involved to create a species count value. Computation of the weighted species count for each cell entry used the following formula:

weighted species count = (species count area / BG area)

These weighted species counts were then tallied to arrive at a total weighted species count for each BG. Because of the variations in a total number of potential species for each vertebrate group, this weighted species count was ineffective in characterizing the relative concentrations of mammals, amphibians, reptiles or

breeding birds. To provide BG-level expected species distribution data, the weighted species count values were converted into percentages of total species expected for each vertebrate class using the following formula (S. Smith et al., 2005):

weighted vertebrate group count = (vertebrate group species count / BG area) / total of potential vertebrate group species

Using percentage of total species enabled comparisons of relative concentrations between vertebrate classes. The methodology described above resulted in an underestimation of species expected within any BG. This underestimation occurs because the species count from any single 30 m^2 grid cell was not adjusted for species composition changes. That is to say, the ten species contributing to the species count for "Cell A" may not be the same ten species contributing to the species count for "Cell B" (S. Smith et al., 2005). For this reason, it is possible that the sum of actual species in Cell A and Cell B could be as few as ten and as many as twenty. Despite this inherent limitation, this approach was deemed acceptable for three reasons. First, the 30 m^2 resolution of the species richness grid was based on the digital satellite imagery used to map plant community types and does not reflect the lower resolution of the vertebrate species data collected (most species data exist at a township or similar scales). Second, since relative species distributions were needed solely to rank the BGs and the underestimation of species was a universal effect, accounting for different species across grid cells within a BG was not required. Third, to account for all possible unique combinations of species would require addressing as many as 2^{214} bird, 2^{25} amphibian, 2^{28} reptile, and 2^{57} mammal combinations.

Table 3.1. Statistical significance of the model variables.

Prediction of Housing Units for 2000					
lm(loghu2000 ~ huden1990 + rd1990den + hu8090 + meandifhuden1990 + poppot1990)					
Variable	Sign	Estimate	StdEr	t value	Pr (>|t|)
intercept		6.14500	0.03112	197.43	<2e -16
huden1990	+	0.00015	0.00002	8.43	<2e -16
rdden1990	-	0.03297	0.00343	-9.62	<2e -16
huchng8090	+	0.00257	0.00017	15.10	<2e -16
poppot1990	-	0.00047	0.00014	-3.47	0.00054
meandifhuden1990	-	0.00006	0.00002	-2.89	0.00388
RSE		0.4242	- 1052 DF		
MR Squared		0.3801		AR Squared	0.3772
F-statistic		129 on 5 and 1052 DF		p-value =	<2.2e-16

Results

All variables are significant. However, the direction of the variable's impact is not always as anticipated. Residual plots of the dependent variable and the predicted fit demonstrate no sign of non-randomness of residuals. The resulting model explained 38% of the model variation. Table 2 presents the impact on the total predicted HUs (housing units) of a 100% change in each predictor variable. This example calculation helps clarify the relative impacts of each of the individual variables on the dependent variable.

Predictions of residential housing unit growth from 2000-2010 were made for each of the 2111 Census 2000 BGs using the above model. From 2000 to 2010, of the 106 BGs predicted to meet the intensive growth area (IGA) threshold, 43 exhibited actual housing unit increases meeting the IGA threshold of placing them in the top 5% of all

BGs for growth (Table 3.2). It was essential to measure the accuracy of the predicted IGAs before flagging conflicts with the vertebrate species concentration areas. Because I had "true" housing unit growth values for the BG geography in the form of Census 2010 data, and the next step of identification of conflicts with species concentration areas requires the overlay of the spatial extents of the "true" IGAs, I selected an error matrix as the evaluation tool for my predictions.

Table 3.2. The impact of an individual predictor variable on the dependent variable.

HU Predicted	Predicted Log HU	=	6.14500 + 0.000015(huden1990) - 0.03297(rdden1990) + 0.00257(hu8090)		
274.394	5.618		- 0.00047(poppo1990) - 0.00006(meandifhuden1990)		
281.315	5.643	--->	change of 100% in huden1990 --->	6.92	change in predicted HU
159.383	5.078	--->	change of 100% in rdden1990 --->	-115.01	change in predicted HU
284.483	5.654	--->	change of 100% in huchng8090 --->	10.09	change in predicted HU
284.674	5.655	--->	change of 100% in meandifhuden1990 --->	10.28	change in predicted HU
252.304	5.535	--->	change of 100% in poppo1990 --->	-22.09	change in predicted HU

The development of an error matrix, the methodology herein presented, is widely accepted in the remote sensing and land use mapping fields as the standard method used to identify and present site-specific error information (Congalton & Green, 1999, pp. 65-84; Lillesand & Kiefer, 2000, pp. 612-618; Campbell & Wynne, 2011). The error matrix displays correct and incorrect classifications for all categories and overall accuracy. Table 3.3 shows the accuracy assessment matrix for IGA predictions for the 2010 Census.

This accuracy assessment matrix accounts for the classification and misclassification of every BG involved. The predicted classes are on the left side of the matrix and across the top are the actual classes determined from the 2010 Census.

The first row of the matrix identifies (reading from left to right): the number of BGs predicted to be IGAs and actually classified as IGAs by the 2010 Census data (row 1, col 1); the number of BGs predicted to be IGAs, but classified as Non-IGAs by the 2010 Census data (row 1, col b); and the total number of BGs predicted to be IGAs (row 1, col c). The second row of the matrix identifies (reading from left to right): the number of BGs predicted to be Non-IGAs but classified as IGAs by the 2010 Census data (row 2, col a); the number of BGs predicted to be Non-IGAs and classified as

Table 3.3. Predicted intensive growth areas (IGA) accuracy assessment matrix for 2010.

Predicted Intensive Growth Areas (IGA) Accuracy Assessment Matrix

Predicted 2010	Actual 2010: IGA	Actual 2010: Non-IGA	Actual 2010: Total	row
IGA	43	63	106	1
Non-IGA	63	1942	2005	2
Total	106	2005	2111	3
col	a	b	c	

Producer's accuracy (true positive rate) = (1a / 1c) = 40.57%

Overall accuracy (accuracy) = (1a + 2b) / 3c = 94.03%

Overall probability of random agreement = 0.25%

Cohen's kappa = 0.94

Intensive Growth Areas: Positive percent increase in housing units placing the Blockgroup in the top 5% of all Blockgroups.

Non-IGAs by the 2010 Census data (row 2, col b); and the total number of BGs predicted to be Non-IGAs (row 2, col c). The final row of the matrix identifies (reading from left to right): the total number of BGs classified as IGAs by the 2010

Census data (row 3, col a); the number of BGs classified as Non-IGAs by the 2010 Census data (row 3, col b); and the total number of BGs (row 3, col c).

The right side of Table 3.3 shows the calculation of producer's accuracy values, which is comparable to precision or true positive rate. Since the ultimate goal is to identify potential IGA conflicts with species concentration areas, it is more important to correctly predict an IGA than it is for every IGA predicted to be correct. Thus, the producer's accuracy for predicted IGAs is the most important metric to be derived from this analysis.

Based on the 2010 Census, the model correctly identified 40.6% of the BGs that exceeded the residential housing unit growth threshold. The resulting prediction rate of 40.6% is in close agreement with the adjusted RMS of 0.377, returned by the model metrics.

Correct predictions of BGs that will receive growth offers a solution to only half of the problem. The other concern is how many of these BGs are coincident with areas of high species diversity or unique species assemblages. Using the weighted vertebrate group concentration values computed for each BG, the areas of intersection of high species diversity and predicted high growth were determined.

The results of the intersection of species richness areas with intensive growth areas predicted and actual are shown in Table 3.4 and indicate that of the 106 BGs identified as the actual IGAs from the 2010 Census, five (5) also received species richness classification and the model correctly identified two (2) of those BGs. In other words, 40% of the species richness BGs that received intensive growth were correctly identified by the model.

Data presented in Table 3.4 also illustrates that of the 211 BGs identified as possessing species richness (SRA), 206 did not receive substantial housing unit growth in the 1990-2000 decade. Of the SRAs that did not meet the IGA threshold, 97% were among those predicted to be below the threshold. This information substantially reduces the focus of concern if conservation of species is a primary goal.

Table 3.4. Predicted species richness areas coincident with predicted intensive growth areas.

	SRA	Non-SRA	Total
IGA Predicted	8	98	106
IGA Actual	5	101	106
IGA Predicted+Actual	2	41	43
Non-IGA Predicted	203	1802	2005
Non-IGA Actual	206	1799	2005
Non-IGA Predicted+Actual	200	1742	1942

IGA (Intensive Growth Areas): Positive percent increase in HU placing the BG in the top 5% of all BGs.

SRA (Species Richness Areas): BGs with species counts placing them in the top 10% of all Block Groups.

Discussion

The results support my hypothesis that residential growth occurs in patterns capable of being modeled at regional scales using variables derived from Census and existing infrastructure data. Modeling these growth patterns will enable the prioritization of at-risk biodiversity areas as defined by the total of all expected species (birds, amphibians, reptiles, and mammals). The model presented produces predictions

of Census BGs likely to receive substantial growth in the next decade with an average accuracy rate of 40%.

When the IGAs, identified by the model, are subsequently intersected with the species concentration areas, they identify a substantially reduced group of species richness areas at risk of impact resulting from future residential development. This winnowing process reduced the group of species richness areas by 96% while still correctly identifying 40% of those at risk.

This information could have reliably focused concern for the conservation of species diversity to the 106 predicted BGs representing 5% of the 2111 BGs in the area, flagging eight species richness areas for concern. Two of these eight BGs (25%), were actual locations of species richness and IGA from 2000-2010.

Utilizing this model, a state agency or regional management entity can substantially reduce their costs and efforts expended to conserve the biodiversity of a region, as defined by total species present, with a reasonable expectation of success. Because this model uses publicly available data, it is easily transportable and can benefit other locales.

It is evident that the concept of predicting residential growth, as represented by housing units, at a regional scale is possible, although a more robust model is desirable. Only one variable in this model contained an overtly spatial component, and yet residential development is all about location.

Improvements to the model must include substantially more spatial data. The inclusion of new inter-BG interactions and intra-BG effects is desirable.

Future model development will be towards dasymetric mapping of the BG characteristics to a grid cell format scaled to the smallest BG in the study area. This type of mapping will involve creating restrictive layers for the study area, identifying grid cells appropriate or inappropriate for residential development based on environmental conditions and political restrictions. These conditions could then be aggregated to obtain a single ranking for each grid cell in the study area. The BG geography will then be overlain and BG characteristics apportioned to the grid cells based on the rankings and land use/land cover data (Sleeter and Gould, 2007; Holt et al., 2004; Eicher & Brewer, 2001). In this way, the BG data will be more accurately distributed on the landscape and raster-based spatial statistical method applied to generate a more robust residential development model.

Mapping of the species concentration areas will also benefit from not being aggregated up to the BG level geography. Using the species data in the 30 m2 cell format provided by the HRV-GAP project, Hotspot Analysis can be implemented on the more accurately delineated species concentrations areas.

Determination of biodiversity areas at risk will also benefit from the inclusion of specific “at risk” species groups such as the “endangered, threatened, and of special concern” groups identified by NYS Department of Environmental Conservation (NYSDEC, 1999). The inclusion of at-risk species groups will address the issue that all species are not considered equal. The inclusion of the “endangered, threatened, and of special concern” species will provide a way to prioritize the species concentration areas.

In summary, I have developed a broadly applicable, geospatial residential growth model based on Block Group (BG) Census data that has successfully identified potential intensive residential growth areas (IGA). The model's five BG-based variables: current housing unit density, current road density, previous decade's housing unit change, housing unit density gradient derived from adjacent BGs, and population potential index make it readily transportable to other states and regions. Of the BGs predicted to be areas of fast growth, 53% and 41% were IGAs as computed from 2000 and 2010 Census data, respectively. A spatial intersection of the identified IGAs and SRAs found 16% and 8% match for the 2000 and 2010 Census data respectively. This iterative process reduced the group of species richness areas substantially while still correctly identifying 40% of those SRAs at risk. This information can be used to reliably focus conservation of species diversity efforts to specific regions of concern.

References

Birch, D.L. (1971). Toward a Stage Theory of Urban Growth. *Journal of American Institute of Planners,* 37(2), 78-87.

Burchell, R. W. & Mukherji, S. (2003). Conventional Development Versus Managed Growth: The Costs of Sprawl. *American Journal of Public Health,* 93, 1534-1540.

Campbell, J.B. & Wynne, R.H. (2011). *Introduction to Remote Sensing*. New York, NY: The Guilford Press.

Congalton, R.G. & Green, K. (1999). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, New York, NY: Lewis Publishers.

Eicher, C.L. & Brewer, C.A. (2001). Dasymetric mapping and areal interpolation: implementation and evaluation, *Cartography and Geographic Information Science*, 28(2), 125-138.

GeoLytics. (2000a). *Census 2010 Summary File 1 in 2000 Boundaries*, Geolytics CD-ROM.

GeoLytics. (2000b). *Census 1990 Summary File 1 in 2000 Boundaries,* Geolytics CD-ROM.

Gustafson, E.J., Hammer, R.B., Radeloff, V.C. & Potts, R.S. (2005). The Relationship between Environmental Amenities and Changing Human Settlement Patterns between 1980 and 2000 in the Midwestern USA. *Landscape Ecology*, 20, 773–789.

Hammer, R.B., Stewart, S.I., Winkler, R.L. et al. (2004). Characterizing dynamic spatial and temporal residential density patterns from 1940-1990 across the North Central United States. *Landscape and Urban Planning,* 69, 183-199.

Hawbaker, T.J., Radeloff V.C., Hammer R.B., & Clayton M.K. (2005). Road density and landscape pattern in relation to housing density, and ownership, land cover, and soils. *Landscape Ecology*, 20, 609–625.

Holt, J.B., Lo, C.P., & Hodler, T.W. (2004). Dasymetric Estimation of Population Density and Areal Interpolation of Census Data. *Cartography and Geographic Information Science*, 31(2), 103–121. https://doi.org/10.1559/1523040041649407

Hu, W. (2000, June 7). Industrial Growth Threatens Scenic Hudson River Valley, Group Warns. *New York Times*. Retrieved from http://www.nytimes.com /2000/06/27/ nyregion/ industrial-growth-threatens-scenic-hudson-river-valley-group-warns.html

Lillesand, T.M., & Kiefer R.W. (2000) *Remote Sensing and Image Interpretation*, New Jersey: John Wiley Press.

NYSDEC (1999) *Environmental Conservation Law of New York, Section 11-0535 and 6 NYCRR (New York Code of Rules and Regulations) Part 182*. Albany, NY: New York State Department of Environmental Conservation.

NYSDOS (2001). *Adopting Zoning for the First Time*, Local Government Technical Series, James A. Coon, Albany, NY: New York State Department of State.

Ottensmann, J.R. (1977). Urban Sprawl, Land Values and the Density of Development, *Land Economics*, 53(4) 389-400.

Perkins, G.C, Kutt, A.S, Vanderduys E.P, & Perry J.J. (2013). Evaluating the costs and sampling adequacy of a vertebrate monitoring program. *Aust. Zool.*, 36, 373–380.

Sanchez, T. W. (2000). Land use and growth impacts from highway capacity increases. *Journal of Urban Planning and Development*, 130(2), 75–82. https://doi.org/10.1061/(ASCE)0733-9488(2004)130:2(75)

Sleeter, R., & Gould, M. (2007). Geographic information system software to remodel population data using dasymetric mapping methods. *US Geological Survey, Techniques and Methods*, 11-C2. Reston, VA: USGS. Retrieved from http://pubs.usgs.gov/tm/tm11c2/

Smith, C.R., DeGloria S.D., Richmond M.E., Gregory S.K., Laba M., Smith S.D.,…Weber J.T. (2001a). *A Gap Analysis of New York NY-GAP. Final Report and Data*. U.S. Geological Survey, Biological Resources Division, National Gap Analysis Office, Moscow, ID. 2-Disc CD set (issued October 2001).

Smith, C.R., DeGloria S.D., Richmond M.E., Gregory S.K., Laba M., Smith S.D.,… Hill E.A. (2001b). *An Application of Gap Analysis Procedures to Facilitate Planning for Biodiversity Conservation in the Hudson River Valley, Final Report, Part 1: Gap Analysis of the Hudson River Valley and Part 2: Atlas of Predicted Ranges for Terrestrial Vertebrates in the Hudson River Valley*. New York Cooperative Fish and Wildlife Research Unit. Ithaca, NY:Department of Natural Resources, Cornell University.

Smith S.D., Brown W.A., Smith C.R., & Richmond M.E. (2004). *Habitat Vulnerability Assessment for the Hudson River Valley: Final Report.* New York Cooperative Fish and Wildlife Research Unit. Ithaca, NY:Department of Natural Resources, Cornell University.

Smith S.D., Brown W.A., Smith C.R., & Richmond M.E. (2005). *Habitat Vulnerability Assessment in the Hudson River Valley*, GAP Analysis Bulletin No. 13, December, 37-43.

Theobald, D.M., Hobbs, N.T., Bearly, T. et al. (2000). Incorporating Biological Information in Local Land-Use Decision Making: Designing a System for Conservation Planning, *Landscape Ecology*, 15:35-45.

Theobald, D.M. (2001). Land-Use Dynamics Beyond the American Urban Fringe. *The Geographical Review,* 91 (3): 544-564.

Theobald, D. M., & Hobbs, N. T. (2002). A framework for evaluating land use planning alternatives: Protecting biodiversity on private land. *Ecology and Society*, *6*(1).

USCB (1994). Geography Division New York State. TIGER/Line Files. Updates to 1990, U.S. Census Bureau, Washington, DC.

USCB (1999). Housing Unit Review File. New York State. Population Estimates Branch, U.S. Census Bureau, Washington, DC.

USCB (2000). Census 2000 Summary File 3A, New York, U.S. Census Bureau, Washington, DC.

USCB (2002). Census 2000 Summary File 3. Technical Documentation. U.S. Census Bureau, Washington, DC.

White, D., Minotti, P.G., Barczak, M.J. et al. (1997). Assessing Risks to Biodiversity from Future Landscape Change, *Conservation Biology*, 11(2), April 1997, 349-360.

CHAPTER 4:

IMPACT OF DASYMETRIC MAPPING ON A CENSUS-BASED REGIONAL SCALE RESIDENTIAL GROWTH MODEL

Abstract

The effects of urban sprawl and the threats thus incurred by ecologically sensitive areas have been widely documented, and the concern of the effects of sprawl has stimulated attempts to model urban and suburban growth. This modeling of land urbanization must integrate the modeling techniques of a social scientist with those of the physical scientist to address both the human and landscape scale variables. Because many social variables are collected by the US Bureau of the Census using their statistical geography of Tracts, Block Groups, and Blocks within the governmental geographies of County, State, and Nation, the issue of the modifiable areal unit problem arises. Can aggregate data presented in one of these geographies be spatially allocated accurately to sub-areas within the geography?

This paper reports on an effort to improve an earlier residential housing prediction model based on regression analysis of Census-based Block Group data and physiographic variables aggregated to the BG level geography. This improvement was sought through the use of dasymetric mapping techniques. Dasymetric mapping is a methodology which seeks to improve the spatial representation and distribution of data values generalized into the areal unit.

The National Land Cover Datasets (NLCD) for 2001 provided a means to sub-divide the Census 2000 BG for allocation of housing unit data based on the NLCD

classification and additional data. Use of alternative Census-based datasets such as the National Historic GIS Time Series and GeoLytics Normalized datasets are discussed.

Introduction

The effects of urban sprawl and the threats thus incurred by our ecologically sensitive areas have been widely documented (Turner et al., 2003; Burchnell & Mukherji, 2003; Su et al. 2012). To quantify the effects of sprawl, Burchnell and Mukherji (2003) modeled county-level growth effects under two development strategies based on national population and employment projections for the period of 2000-2025. Their models provide quantitative results of the effects of sprawl compared to managed growth. Their results provide data on land conversion, water and sewer demand, road construction, and fiscal impacts by national regions (Burchnell & Mukherji, 2003, pp. 1537-1539). In every region of the country projected development costs under sprawl were at least 4 times higher than under the managed growth scenario (Burchnell & Mukherji, 2003, p. 1539).

Driven by the concern of the ecological and fiscal effects of sprawl, attempts to model urban and suburban growth have been stimulated. The modeling of land urbanization has had to integrate the modeling techniques of a social scientist with those of the physical scientist to address both the human and landscape scale variables (Veldkamp, 2004). In addressing this need to merge these modeling techniques, the modifiable areal unit problem (MAUP) identified by Openshaw (as cited in Wong, 2004, p. 571;Dark & Bram, 2007, p. 472;) had to be addressed. The MAUP applies to all data that are derived from surveys of various sampling sizes across the landscape

and aggregated into delineated zones (Dark & Bram, 2007). The issue which arises is that the data aggregation masks the true spatial distribution within the zone of the variable being measured. The user of these aggregated data must assume that the variable value assigned to the zone is evenly distributed throughout the zone. The purpose of dasymetric mapping is the improvement of the spatial representation and distribution of data values generalized into the areal unit (Mennis & Hultgren, 2006). The Census Block Group geography data is derived in this manner. Intensive research has been underway investigating methods of using dasymetric mapping techniques with various parameters and ancillary data to overcome the MAUP and redistribute Census geography data spatially, as well as temporally (Holt et al., 2004; Mennis & Hultgren, 2006; Mitsova et al., 2012).

Dasymetric mapping has utilized binary, weighted, and hybrid techniques to distribute the zonal values to sub-zones designed to more accurately represent the true spatial distribution of the data. That said, this new "sub-zones" still suffer from the MAUP, albeit to a much lesser degree (Holt et al., 2004).

The simple binary approach to dasymetric mapping takes the given zonal area and divides it into sub-zones based on a binary classification (Langford, 2007). In the case of Census data, it is often inhabitable vs uninhabitable areas (e.g., land vs water) as identified by ancillary sources (Holt et al., 2004). Variable values are assigned to the sub-zones based on area weights (Eicher & Brewer, 2001). In other words, if a habitable sub-zone is one fourth the area of all habitable sub-zones within the zone then it would be assigned 25% of the zone's population.

A modification to this method which can increase the accuracy of the variable distribution is the further subdivision of the binary sub-zones based on another ancillary dataset such as land use / land cover (LULC). This three variable method could use a generalized LULC classification layer, intersect it with the habitable and uninhabitable binary sub-zones to further divide the zones, and weight the assignment of the population by the LULC classification (Mennis, 2003). This method requires the designation of weights for each LULC class mapped. These weights may be derived subjectively or from empirical sampling.

Hybrid models of dasymetric mapping add additional layers to augment or modify other layers to improve a layer which may have been generated at a spatial or taxonomic scale too generalized for the study area. For example, the addition of county tax parcel data is often used to refine the classifications of satellite and NLCD datasets (Sleeter & Gould, 2007; Mitsova et al., 2012; Jia & Gaughan, 2016; Zoraghein et al., 2016).

With the goal of empowering local officials, I am attempting to predict the spatial distribution of the residential expansion through the use of readily accessible public data. My earlier models, based on regression analysis of Census-based and physiographic variables aggregated to Census Block Group (BG) level geography (see Chapters 2 and 3), have provided reasonable predictive power at the Census BG level (USCB, 1994).

What I am addressing in this chapter is whether improvements to the previous model can be obtained through the application of dasymetric mapping of Census BG data into spatial subsets. The ability to predict, at a regional scale, probable residential

growth areas and to overlay those identified areas with locations of species richness, will highlight high biodiversity areas at risk. The Census data provides decadal-series data enhancing the model's temporal portability. Developing the species richness areas from the National GAP Analysis Program, a national inventory of vertebrate species, combines two country-wide datasets providing spatial portability (Smith et al., 2001). Having the ability to focus regulatory and educational efforts to locales, identified in advance, of probable residential growth that also support a high number of vertebrate species enables resource managers and conservation groups to improve the efficiency of their efforts and monetary expenditures.

Based on the earlier model successes, I hypothesize that the spatial accuracy of the existing residential growth model can be improved using dasymetric methods to distribute BG data to spatial subsets while maintaining the effectiveness of the model to identify at-risk biodiversity areas. This study has the following objectives: 1) Determine an appropriate method for implementation of dasymetric mapping to the BG variables utilized in the previous model, 2) Determine an appropriate spatial resolution for the dasymetric results, and 3) Implement the previous model at the new spatial resolution with the dasymetrically modified BG variables.

Study area

The adverse effects of sprawling urbanization on ecologically sensitive areas has prompted a growing effort to understand spatial patterns of residential development. As with my earlier models, the study area is the ten counties in the Hudson River Valley (HRV) area. Specifically, the counties of Albany, Columbia, Dutchess, Greene,

Orange, Putnam, Rensselaer, Rockland, Ulster, and Westchester (Figure 4.1). A significant reason for selecting the HRV was the amount of physiographic, anthropogenic, and biological diversity represented.

The biological diversity of the HRV has been detailed in earlier chapters. To provide a different perspective on the diversity present in the study area, this chapter will focus on specific physiographic and anthropogenic characteristics of the counties involved. The diversity of the area's anthropogenic and physiographic characteristics is graphically displayed by county in Figures 4.2 and 4.3, respectively.

Figure 4.2 displays the selected anthropogenic characteristics: percent of total study area, road length, number of municipalities, population, and developable land for each county in the study area.

A substantial county-to-county diversity is evident with five counties (Albany, Dutchess, Orange, Putnam, and Rensselaer) having internal percentage values that vary relatively little, less than 5%, between characteristics while the remaining five counties provide the intra-county diversity (Table 4.2).

Ulster easily stands out as the largest county, occupying almost 19% of the total study area. Six counties (Albany, Columbia, Dutchess, Greene, Orange, and Rensselaer) have land areas, presented as a percentage of the total study area, within five percent of each other and occupy 67% of the region (Table 4.2).

Road length is an important indicator of existing development intensity and potential for further development. A higher density of road development is indicated by the percentage of road length to land area. Westchester, Rockland, and to a lesser extent, Putnam Counties exhibit this higher road length to land area ratio (Table 4.2).

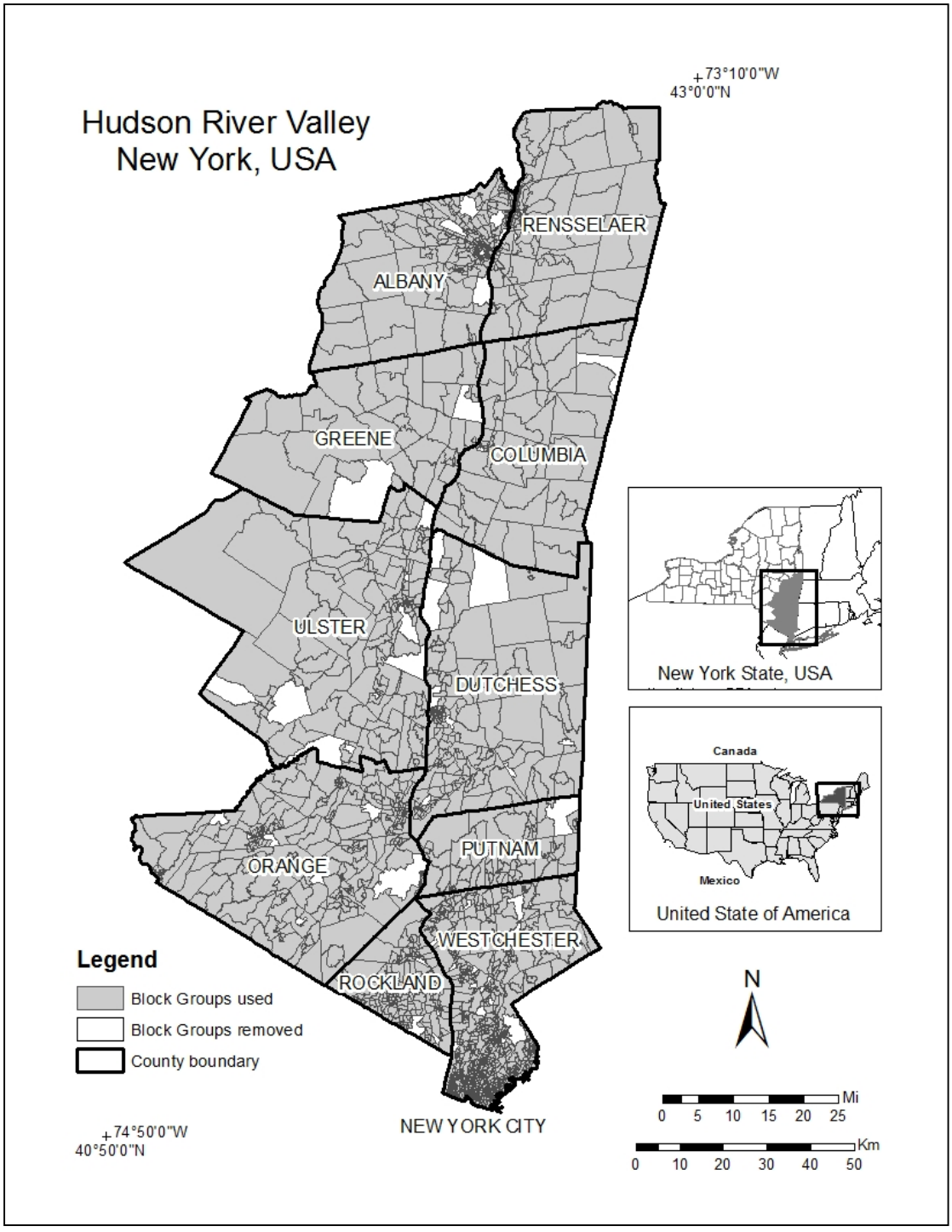


Figure 4.1. Ten-county region with the 2,111 Block Groups used in the analysis and the 101 omitted Block Groups that represented group quarters and institutional housing. Group quarters and institutional housing increases are subject to institutional and governmental forces outside the scope of this study.

A higher amount of existing development is expected to accompany this higher road density. Counties such as Greene, Ulster, and Rensselaer exhibit the flip of the ratio and suggests land area available for new housing development and associated roads (Tale 4.2).

Population percentages, developed from the 2000 Census (USBC, 2000), are best compared with land area to yield population density. The lowest density counties (Columbia, Greene, and Ulster) are rural counties having an average population/land area ratio of 0.90 and in contrast, the two developed counties of Rockland and Westchester exhibit a 3.96 population/land area ratio (Table 4.2). Barring impacts from other variables, counties with lower population densities should be more capable of supporting increased development than counties with higher population densities.

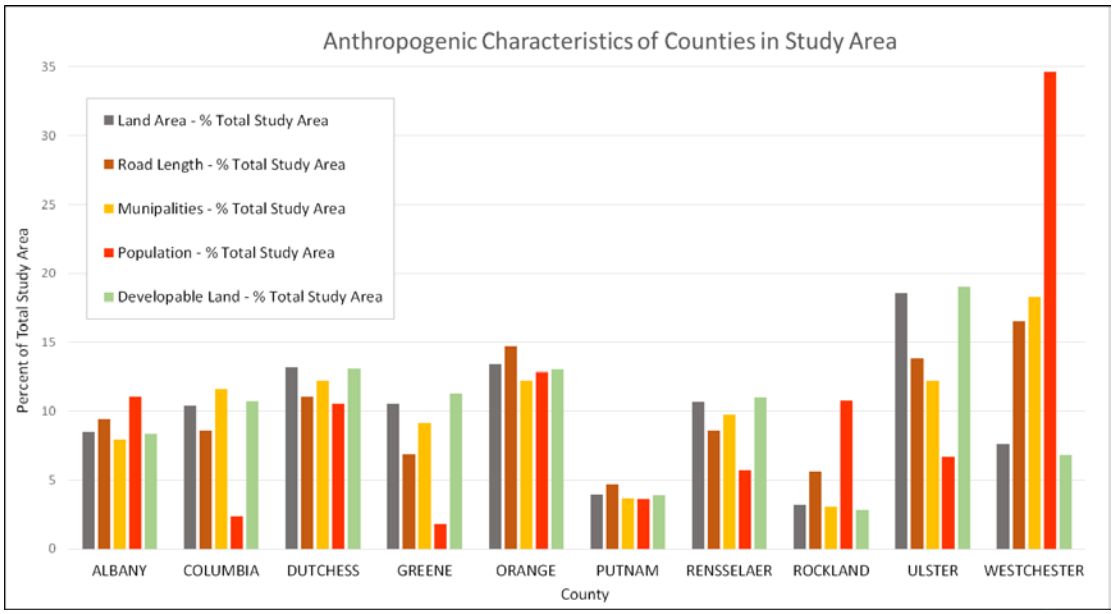


Figure 4.2. Selected anthropogenic characteristics of each county in the study area as a percentage of total study area characteristics.

I have defined developable land as those areas: 1) outside governmental restrictions as identified as Status 1-3 by the NY-GAP project (Smith et al., 2001) and 2) not within the stream buffers, open water, wetland or steep slope areas. Ulster County's high percentage of the developable land, coupled with low population density, results in a population density to the developable land ratio of 0.35, suggesting a higher potential for development (Table 4.2).

Figure 4.3 shows the following natural characteristics of each county as a percentage of the total study area: land area, steep slope, surface water, river shoreline, and wetlands. There is a substantial variation of percentages within the counties as well as between the counties. Land area distribution is the same as in Figure 4.2 and is provided for comparison purposes.

Steep slopes are defined here as slopes with gradients of 25% or greater. These steep slope areas present challenges to standard building practices and increase development costs. Steep slopes are distributed in Greene and Ulster Counties, 20 and 35% respectively, largely due to their inclusion of portions of the Catskills Mountains (Table 4.3).

Surface water within the study area includes major streams, lakes, and the Hudson River. Proximity to water is a development feature which possesses positive opportunities and negative challenges. Individuals are attracted to water for the visual and emotional impacts as well as recreational activities. Development of land highly fragmented by water can present siting problems, engineering concerns and, increased costs.

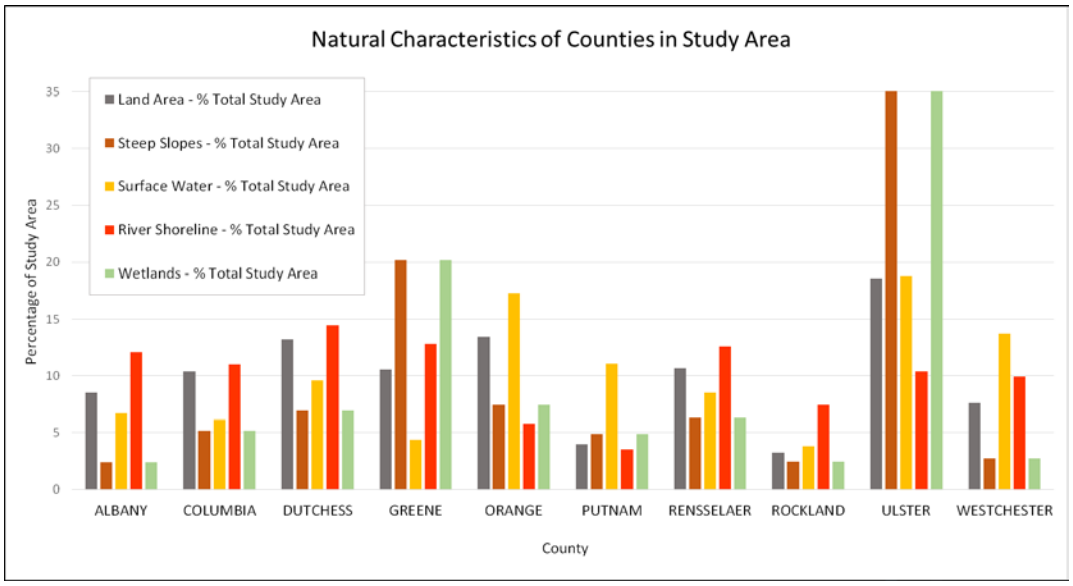


Figure 4.3. Selected natural characteristics for each county in the study area as a percentage of total study area characteristics.

Every county also includes Hudson River or Mohawk River shoreline, which is shown as a separate factor in Table 4.3 because it provides opportunities uniquely different than other surface water features. Both scenic rivers are navigable and under regulation by federal (USACE, 2017) and state (NYDEC, 1999) agencies.

The final natural characteristic presented in Table 4.3 is the percentage of wetlands possessed by each county. The wetlands measured are from the National Wetlands Inventory which has mapped wetlands that are regulated (USFWS, 2000).

The importance of this study area is enhanced by the amount of biodiversity it contains, as defined by supported terrestrial vertebrate species. The most recent regional gap analysis project found this ten county area to be inhabited by 80% of the resident terrestrial vertebrate species in New York State (Smith et al., 2001).

Methods

My earlier model used the Census BG geography as the unit of analysis based on the statistical nature of BG boundaries and the availability of a reference boundary. Optimally, Census BGs contain approximately 400 housing units. The actual counts may vary from 250 – 550 housing units. (USBC, 1994). Although the size and boundary delineation of a BG may change between decadal census, the Census Bureau has developed a methodology for relating the BG data from one census period to another. Census data for the 1990 and 2010 surveys, normalized to the Census 2000 boundaries, were prepared using the Census Bureau's methodology and made available by a private company, GeoLytics, Inc. (GeoLytics, 2010). This normalization of decadal data was instrumental in the effectiveness of the earlier model as documented in Chapter 3.

To improve the model's predictive power, I sought to improve the spatial resolution of the data aggregated to the Census BG geography. The desire to improve the spatial scale of the BG data involves addressing the MAUP (Dark & Bram, 2007). Because of this need, I pursued the application of dasymetric mapping techniques to the decadal census data.

Drawing upon the success of efforts employing hybrid dasymetric mapping methods (Eicher & Brewer, 2001; Holt et al., 2004; Mitsova et al., 2012), I developed the methodology used in this study.

I have undertaken the effort to implement the dasymetric mapping of Block Group data to smaller polygons defined by the intersection of Census BG geography, a derived non-developable area layer, and National Land Cover Database (NLCD)

polygons (Homer, 2007; USGS, 2014). Although the successful application of this methodology included the use of ancillary tax parcel boundary data (Sleeter & Gould, 2007), I had to forego the use of the parcel data because it was not available for all counties within the study area.

The model for this effort used the NLCD 2001 (Homer, 2007; USGS, 2014) land use data to distribute the Census 2000 BG housing units (HU) onto the landscape using GeoLytics Census products. The GeoLytics products provided Census tabulations for 1990, 2000, and 2010 decadal surveys adjusted to the Census 2000 BG boundaries. This product, used in my early models, enabled analysis across the decadal surveys.

The regression equation used in an earlier study (Chapter 3) required five variables to predict the log of the housing units count for 2010. These five variables, housing unit density, road density, housing unit change from the previous decade, population potential, and the difference between the housing unit density of a BG and the mean housing unit density of the adjacent BGs are all based on Census 2000 BG geography.

Before creating the BG subsets, the NLCD 2001 layer needed to be modified categorically and spatially to reduce the total number of subsets created. The first step was to reclassify the layer from the 15 NLCD 2001 classes found in New York State to a total of five (5) classes addressing the needs of the model. Figure 4.1 contains the class aggregations undertaken to obtain the following five classes: rural, low-density residential, medium-density residential, high-density residential/commercial/industrial, and non-developable. Since the model's focus is the identification of future growth areas, the classification scheme combines high-density residential with the other

highly impervious surface classes. This consolidation was supported by the premise that existing high-density residential areas have already delivered their impacts upon the habitats and are close to being fully developed, thereby possessing a low potential for further substantial residential development. My primary interest is in the areas that are capable of supporting and likely to attract substantial additional residential development.

To improve the spatial accuracy of my housing unit growth predictions, I generated a non-developable layer for the study area. This non-developable layer is based on physical, legal, and governmental constraints on residential development. This non-developable layer is a union of the following datasets: surface hydrography from New York State Department of Environmental Conservation (NYSDEC, 2013) with linear features buffered by 30 meters each side; federal regulated wetlands from the National Wetlands Inventory (NWI); slopes greater than 25% from NYS DEC (USGS, 1997) 10 meter digital elevation model; and areas at a stewardship level of 3 and above (e.g. national and state parks and other protected areas), from the New York GAP Analysis Program (Smith et al., 2001). Any area identified in one of the four contributing restrictive elements is assigned “non-developable” classification. All remaining areas are classified “developable” and are used in the creation of developable subsets within each BG.

The spatial resolution of the NLCD layer, 30 meters, is overly detailed for my purposes. Converting the unmodified NLCD layer to a polygon layer that maintains the 30-meter cell size, as well as the 15 land use classes within New York State results

in more than 1 million polygons. Because I am interested in the identification of sizable areas prone to residential growth that could pose a threat to vertebrate species

Table 4.1. NLCD 2001 classification for land cover mapped in New York State and recoded into five (5) classes.

RECODING OF NLCD 2001 CLASSES USED IN NEW YORK STATE			
NLCD 2001 CODE	DESCRIPTION	5-CLASS CODE	DESCRIPTION
11	Open Water	5	Non-Developable
21	Low Intensity Residential	2	Low Density Residential
22	High Intensity Residential	3	High Density Residential
23	Commercial Industrial Transportat	4	Commercial Industrial Transportation Barren
24	Developed High Intensity	4	Commercial Industrial Transportation Barren
31	Barren Land	4	Commercial Industrial Transportation Barren
41	Deciduous Forest	1	Rural
42	Evergreen Forest	1	Rural
43	Mixed Forest	1	Rural
52	Shrub/Scru	1	Rural
71	Grassland/Herbaceous	1	Rural
81	Pasture/Hay	1	Rural
82	Cultivated Crops	1	Rural
90	Woody Wetlands	5	Non-Developable
95	Emergent Herbaceous Wetlands	5	Non-Developable

habitats, the quarter-acre cell size represented by the 30-meter cell resolution is unnecessarily detailed. The translation of the NLCD 2001 layer into a 180-meter (3.25 hectares or 8 acres) resolution dataset resulted in a higher spatial resolution than 99% of the Census BGs and ultimately reduced the total number of polygons processed to a reasonable 18,000. To accomplish this translation, I resampled the 30-meter NLCD 2001 dataset into a 180-meter dataset using a majority filter based on the surrounding eight cells. This aggregation is accomplished in two stages. First, the 30-meter data were aggregated with a majority filter to result in a 90-meter resolution data. This 90-

meter resolution data was then passed through a majority filter with the resulting output being 180-meter resolution data. Although the 180m cell resolution would mask some of the variation in the urbanized areas, this loss of detail was deemed acceptable given the goal of identifying areas with the potential to exhibit substantial housing unit growth. A 180m x 180m cell resolution is an area of 32,400 m^2 which is the equivalent of 3.24 hectares (8 acres). It was also thought that using a 180m cell resolution might compensate for the spatial inaccuracies evident in some of the Census 2000 BG boundary delineations.

The resulting 180m x 180m, five class NLCD 2001 modified dataset was intersected with the Census 2000 BG boundary layer to produce BG-subsets (16,147 polygons) based on land cover characteristics. The process of allocating the BG characteristics to these polygons required the determination of appropriate housing unit densities to associate with each of the five land cover classes. To determine the densities for each land cover class, I turned to the Census 2000 Block data. Blocks are the smallest Census geography containing population and housing counts. I intersected the Census 2000 Block layer with the original 30-meter NLCD data layer and extracted NLCD class and Block area for each Block comprised of a single land cover class. The Census 2000 housing unit densities for these selected BGs were plotted to acquire median housing unit density values with upper and lower ranges for each of the four land cover classes offering potential areas for new housing (i.e., excluding non-developable land cover class). A weighting scheme (Figure 4.4) was developed to proportionally assign housing unit values to each NLCD subset area within a BG.

The process of assigning housing unit densities begins with the determination of median HU densities per land cover class derived from the single NLCD class Census Blocks: class 1 = 2.5; class 2 = 5.5, class 3 = 13.3; and class4 = 24.9 (Figure 4.4, Step 1). The area for each land use class is then normalized to the total developable area within the BG to yield land class weight (Figure 4.4, Step 2). The median HU densities per land cover class were also normalized based on the total of all possible median HU densities found within the BG (Figure 4.4, Step 3) to derive the BG HU weight. A BG containing all four developable classes (rural, low density, medium density, and high density) would use the value 46.2 for normalization, the sum of class 1 (2.5), class 2 (5.5), class 3 (13.3) and class 4 (24.9).

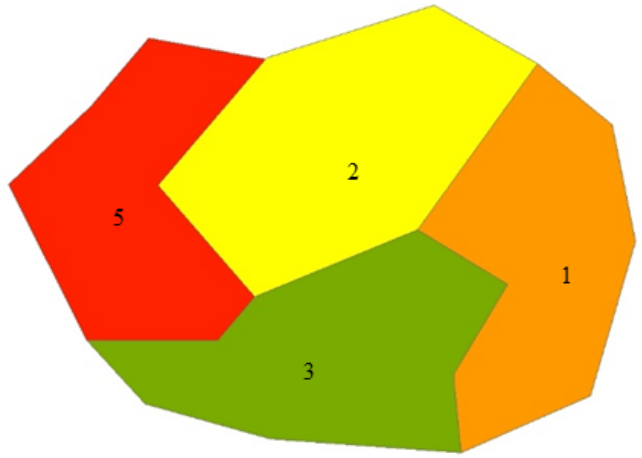
The product of the land class weight and the BG HU weight yielded the land class BG HU weight value (Figure 4.4, Step 4). This land class BG HU weight value is then normalized to the sum of all land class BG HU weight values to provide the normalized land class BG HU weight (Figure 4.4, Step 5). Taking the normalized land class BG HU weight and multiplying it by the total HU for the BG yields the portion of HU to be allocated to a specific land class (Figure 4.4, Step 6).

This total housing unit value is then divided by the total area of land cover within the BG to yield the housing unit density per hectare for those BG subsets. This same procedure was also used to allocate housing unit change in the previous decade to the BG land cover subsets. To ensure that all BG HU's were accounted for, the above methodology was check by regenerating the original BG HU values from the subset values (Tobler, 1979; Kim & Yao, 2010).

The other three independent variables required regeneration for each BG land cover subset: road density; population potential; and the difference between the housing unit density and the mean housing unit density of the adjacent BGs.

BG Assumptions
BG Area = 100 ha
Restricted area = 20 ha
Developable area = 80 ha
BG Hu total = 45

NLCD Class Assumptions
Class 1 20 ha
Class 2 25 ha
Class 3 35 ha
Class 4 0 ha
Class 5 20 ha



NLCD Class	Step 1 Median HU Per Land Class	Step 2 Land Class Weight	Step 3 BG HU Weight	Step4 Land Class Weight * BG HU Weight	Step5 Normalized (Land Class Weight * BG HU Weight)	Step6 HU Per Class Area
Class 1	2.500	0.250	0.054	0.014	0.077	3.446
Class 2	5.500	0.313	0.119	0.037	0.211	9.475
Class 3	13.300	0.438	0.288	0.126	0.713	32.079
Class 4	24.900	0.000	0.000	0.000	0.000	0.000
	46.200			0.177		45.000

Figure 4.4. Weighting schema for proportional assignment of housing units to BG land cover subsets. BG and NLCD class assumptions provided for illustrative purposes only. Step 1 - determine median HU densities per land cover class; Step 2 - normalize land class area are normalized to total developable area within the BG; Step 3 - median HU densities per land class weights are normalized to the sum of all possible median densities within the BG; Step 4 – determine product of land class weight and the BG HU weight; Step 5 – normalized the product in Step 4 to the sum of all products in Step 4; and Step 6 – multiply the normalized product from Step 5 times the total BG HU to yeild the total HU allocation for each land class within the BG.

Road density was calculated for each BG-subset excluding the area contained with an internal 30-meter buffer of the BG's boundary and inside any non-developable area in the BG to remove the influence of roads that were shared with other BG's and concentrated on the internal road density of the BG's developable area.

Population potential was calculated for each BG-subset created by the intersection of the modified NLCD 2000 layer and the BG boundaries. Distance to each of the 20 population centers identified in the previous study was calculated using the centroid of the BG-subsets. Since the population potential variable uses population rather than housing units, the BG's population was allocated to each BG-subset based on the ratio of allocated housing units allocated to each BG-subset and the BG's total housing units. This allocation is based on the assumption that within a BG, the population is directly related to housing units.

The third variable, mean housing unit density gradient, is the difference between the housing unit density and the mean housing unit density of the adjacent BGs. This variable utilizes the same BG-subsets as the model geography. The mean housing unit density of the adjacent BG-subsets (source sum) was calculated for each BG-subset (target) and the difference (source sum - target) yields the gradient. A positive gradient would imply pressure for growth from adjacent BG-subsets and conversely a lack of pressure for growth would be implied by a negative gradient.

Once all the variables needed for the model were calculated for each BG-subset, I imported the data into the R statistical package and applied to the model below.

loghu2010 ~

huden2000 + roadden2000 + huchng9000 + poppot2000 + meandifhuden2000

Results

Implementation of the model with Census 2000 BG boundary data, the NHGIS 2010, and the Geolytics 2010 dataset produced poor results (Table 4.2). These results highlight the difficulties confronted when trying to implement a dasymetric mapping solution to address the MAUP. Faced with unsatisfactory results, I investigated possible causes as to why the efforts to improve spatial accuracy failed. Three potential causes were identified: the aggregation of the NLCD to a 180-meter cell resolution, the NLCD land cover classification, or the Geolytics BG layer data.

Although the 180-meter cell aggregation appeared to retain the important LULC classification and represented an area wherein reasonable housing development could occur, a smaller resolution may be advantageous. A review of the NLCD land cover layer revealed that the low-density areas very often included golf courses and cemeteries, neither of which would contribute to increased residential development (Figure 4.5). This is not an interpretation error but an error in the classification scheme with respect to my usage. Since no alternative to using the NLCD layer existed and interpreting a regional land cover layer would involve substantial time and violate a primary goal, use of existing digital data, I sought an alternative. Other studies have utilized tax parcel data to modify satellite imagery data (Sleeter & Gould, 2007; Mitsova et al., 2012), but since the parcel boundary dataset was not available for the whole study area, one option would be to perform an analysis on a subset of the study area. Before moving in that direction, I investigated other alternatives.

Table 4.2. Prediction of Housing Units for 2010 using dasymetric mapping methodology.

Prediction of Housing Units for 2010					
lm(loghu2010 ~ huden2000 + rd2000den + hu9000 + meandifhuden2000 + poppot2000)					
Variable	Sign	Estimate	StdEr	t value	Pr (>|t|)
intercept		20.63000	0.32400	63.67	<2e-16
huden2000	+	81.36000	20.22000	4.02	6.16e-05
rdden2000	-	73.12000	49.74000	-1.47	0.142
huchng9000	+	0.03118	0.00288	10.83	<2e -16
meandifhuden2000	-	0.00365	0.00356	-1.03	0.30600
poppot2000	-	0.00002	0.00108	0.01	0.98800
RSE		4.323	- 1003 DF		
MR Squared		0.1214		AR Squared	0.117
F-statistic		27.71 on 5 and 1003 DF		p-value =	<2.2e-16

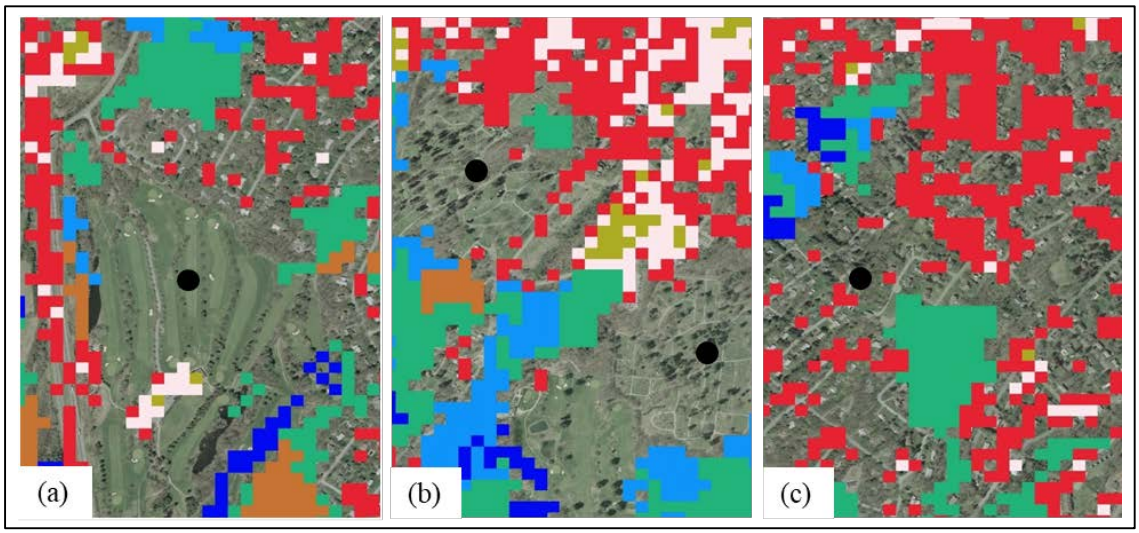


Figure 4.5. Example of NLCD classification of golf courses (a), cemeteries (b), and low-density development (c) areas identified by black dots. The NLCD class of low-density development (class 21) is set as tranparent to enable viewing of 2004 aerial imagery displayed underneath.

A review of the Geolytics BG boundaries for Census 2000 dataset determined that some of the boundaries are substantially misaligned when viewed with newer imagery and authoritative base maps (Figure 4.6). This misalignment occurs sporadically throughout the region. Correcting these boundaries would rely upon the interpretation of imagery, base maps, and an assumption of where the lines belonged or access to original documents. A concern was that adjusting the BG boundaries might introduce more error into the dataset.

To move forward with the investigation of whether dasymetric mapping of Census BG data will improve model results, an alternative direction was needed. The Minnesota Population Center (MPC) at the University of Minnesota developed a dataset which may provide the solution. As part of the National Historical GIS (NHGIS), the MPC developed a time-series dataset which allocates Census 2000 data values into Census 2010 block group boundaries (Manson et al., 2017). In the process of BG data translation, MPC uses a newly developed hybrid allocation method to translate the data. This hybrid method uses the Census Bureau relationship files in conjunction with dasymetric mapping techniques to distribute the Census 2000 BG data into Census 2010 BG geography (Schroeder, 2017). This methodology improves upon the method relying solely on the Census relation files by including the use of ancillary data into the proportioning process.

The advantage of the NHGIS time series dataset is the use of “4 types of ancillary data to guide interpolation: 2010 block densities, imperviousness data, road buffers, and water body polygons” (Schroeder 2017, p. 53). These ancillary data fuel a hybrid

model based on target-density weighting and binary dasymetric models BD. In the majority of cases the results varied little from the non-hybrid model.

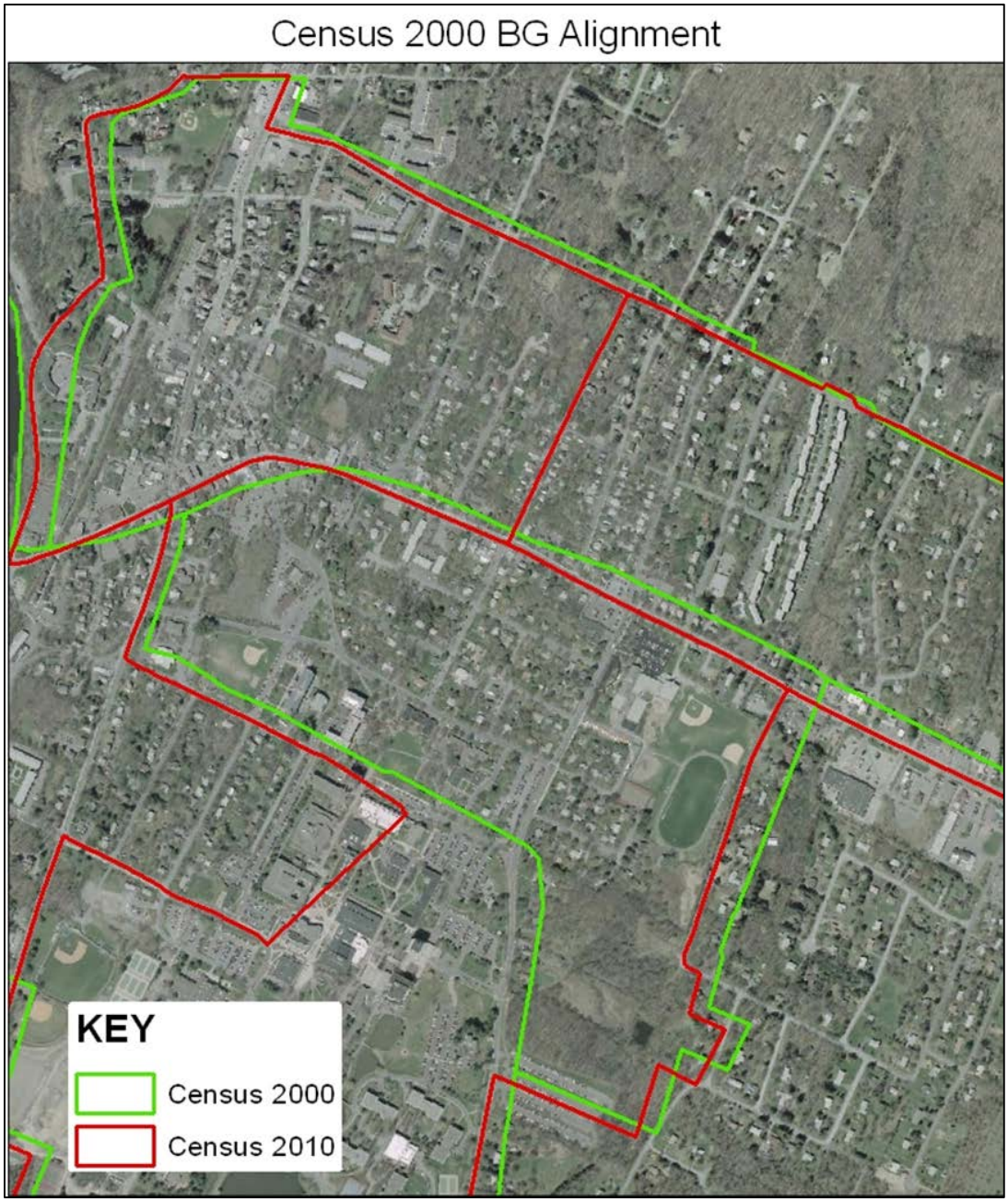


Figure 4.6. Alignment of Census 2000 Block Group boundaries compared to Census 2010 Block Group boundaries and orthoimagery.

Using the NHGIS time series dataset would move all of the BG data into the Census 2010 BG boundary geography, improving the boundary alignments since the 2010 BG delineations aligned more precisely with road and land cover datasets. Reprocessing the data using variable values from the NHGIS time series dataset (Manson et al., 2017) did not significantly improve the model results.

Continuing with the investigation of how to improve the results, I decided to check the NHGIS time series results against a similar dataset prepared by GeoLytics, Inc. I acquired the GeoLytics normalized dataset (Census 2000 in 2010 Boundaries) which was also based on the Census 2010 BG delineations (GeoLytics, 2010). A major difference between the GeoLytics 2010 and the NHGIS 2010 datasets is the addition of the hybrid dasymetric allocation process used to create the NHGIS time series dataset. Reprocessing the data using variable values from the GeoLytics 2010 dataset did not significantly improve the model results.

A final issue to consider is that the shift to using Census 2000 data mapped into Census 2010 geographies may be affecting the underlying variable relationships upon which the earlier models relied. The foundational relationships may have changed in the process of distributing the data via the GeoLytics and the NHGIS methodologies. To test whether the variable relationships had been altered, the GeoLytics 2010 and the NHGIS 2000 time series BGs were split into two random groups to train and test the model original BG level model. The model was able to achieve 37% RMS when applied to the NHGIS 2010 dataset using BG level geography. Applying the original BG level to the GeoLytics 2010 dataset provided similar results.

Discussion

Although the results experienced in this use of dasymetric mapping of Census BG data for implementation of the model are less than satisfactory, they have provided numerous questions for further investigations, such as the following:

1. Does this effort suggest that my earlier model may not have been robust enough or perhaps tailored to the dataset?

All indications are that my earlier model was correctly developed and tested. The unsatisfactory results using dasymetric mapping could be an issue with the data, my manipulation of the data, or perhaps a change in the development drivers since the last decade.

2. Was my earlier model impacted by the misalignment of the Census 2000 BG boundaries with the ancillary datasets?

It would seem very likely that my model was impacted to some degree by the misalignment of the Census 2000 BG boundaries. However, I do not believe that the alignment issue had a substantial impact on the variables used in the non-dasymetric models. This view is supported by two factors. First, the only variable used in the non-dasymetric model that was tied directly to another dataset was road density. A study of road density values before and after Census 2000 BG boundary corrections may help to quantify the level of the impact. Secondly, the testing of the non-dasymetric model using the GeoLytics 2010 dataset at BG level produced almost identical results as the GeoLytics 2000 dataset based on the misaligned BG boundaries, suggesting that the BG boundary alignments were not an issue.

3. Could an implementation of the dasymetric model be successful after improving the NLCD classifications impacting rural and low-density areas?

The attractiveness of this direction is that I can retain all of the current variables within a sub-BG polygon structure and process the regression on those polygons. I also believe it to be prudent to reduce the aggregation of the original 30-meter resolution NLCD data to a maximum of 90-meter resolution. The spatial aggregation of the NLCD 2001dataset to the 180-meter resolution may have overly generalized the data.

4. Would an implementation of a dasymetric mapping methodology using an NLCD 2001 dataset modified using property parcel centroid information and then aggregated into 90-meter cells to create sub-BG polygons provide a pathway to a successful model?

This certainly may result in improvements to the model's effectiveness and is worth investigating, but the possibility exists that my model is not transferable to smaller sampling areas than BGs. My original goal was the development of an easily implemented regional model that is portable and I successfully produced such a model. The difficulties that have arisen are the result of trying to scale the model down to sub-block group areas. I may need to face the possibility that the model cannot be implemented successfully at that scale and investigate ways to increase the model's robustness at the BG or similar spatial level.

References

Burchell, R. W. & Mukherji, S. (2003). Conventional Development Versus Managed Growth: The Costs of Sprawl. *American Journal of Public Health*, 93, 1534-1540.

Dark, S. J., & Bram, D. (2007). The modifiable areal unit problem (MAUP) in physical geography. *Progress in Physical Geography*, *31*(5), 471–479. https://doi.org/10.1177/0309133307083294

Eicher, C.L. & Brewer, C.A. (2001). Dasymetric mapping and areal interpolation: implementation and evaluation, *Cartography and Geographic Information Science*, 28(2), 125-138.

GeoLytics. (2010). Census Data 2000 in 2010 Boundaries, GeoLytics, Inc., Retrieved from http://www.geolytics.com/USCensus,Normalized-Data,Categories.asp

Holt, J.B., Lo, C.P., & Hodler, T.W. (2004). Dasymetric Estimation of Population Density and Areal Interpolation of Census Data. *Cartography and Geographic Information Science*, 31(2), 103–121. https://doi.org/10.1559/1523040041649407

Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C.,…Wickham, J. (2007). Completion of the 2001 National Land Cover Database for the Conterminous United States. *Photogrammetric Engineering and Remote Sensing*, 73(4), 337-341.

Jia, P., & Gaughan, A. E. (2016). Dasymetric modeling: A hybrid approach using land cover and tax parcel data for mapping population in Alachua County, Florida. *Applied Geography*, *66*, 100–108. https://doi.org/10.1016/j.apgeog.2015.11.006

Kim, H., & Yao, X. (2010). Pycnophylactic interpolation revisited: Integration with the dasymetric-mapping method. *International Journal of Remote Sensing*, *31*(21), 5657–5671. https://doi.org/10.1080/01431161.2010.496805

Langford, M. (2007). Rapid facilitation of dasymetric-based population interpolation by means of raster pixel maps. *Computers, Environment and Urban Systems*, *31*(1), 19–32. https://doi.org/10.1016/j.compenvurbsys.2005.07.005

Manson S., J. Schroeder, D. Van Riper, & S. Ruggles. (2017). *IPUMS National Historical Geographic Information System: Version 12.0* [Database]. Minneapolis: University of Minnesota. http://doi.org/10.18128/D050.V12.0

Mennis, J. (2003). Generating Surface Models of Population Using Dasymetric Mapping. *Professional Geographer*. https://doi.org/10.1111/0033-0124.10042

Mennis, J., & Hultgren, T. (2006). "Intelligent" Dasymetric Mapping and Its Comparison to Other Areal Interpolation Techniques. *Cartogis.Org*, (215), 1–9. Retrieved from http://www.cartogis.org/publications/autocarto-2006/mennishultgren.pdf

Mitsova, D., Esnard, A. M., & Li, Y. (2012). Using enhanced dasymetric mapping techniques to improve the spatial accuracy of sea level rise vulnerability assessments. *Journal of Coastal Conservation*, *16*(3), 355–372. https://doi.org/10.1007/s11852-012-0206-3

NYSDEC (1999) *Environmental Conservation Law of New York, Section 11-0535 and 6 NYCRR (New York Code of Rules and Regulations) Part 182*. Albany, NY: New York State Department of Environmental Conservation.

NYSDEC (2013). *New York State Regulatory Freshwater Wetlands, 1999-2013*. Albany, NY: NYS Dept. of Environmental Conservation. http://cugir.mannlib.cornell.edu/READMES/fwrdme.html

Schroeder, J. P. (2017). Hybrid areal interpolation of census counts from 2000 blocks to 2010 geographies. *Computers, Environment and Urban Systems* 62, 53-63. http://dx.doi.org/10.1016/j.compenvurbsys.2016.10.001

Sleeter, R., & Gould, M. (2007). Geographic information system software to remodel population data using dasymetric mapping methods. *US Geological Survey, Techniques and Methods,* 11-C2. Reston, VA: USGS. Retrieved from http://pubs.usgs.gov/tm/tm11c2/

Smith, C.R., DeGloria S.D., Richmond M.E., Gregory S.K., Laba M., Smith S.D.,…Weber J.T. (2001). *A Gap Analysis of New York NY-GAP. Final Report and Data*. U.S. Geological Survey, Biological Resources Division, National Gap Analysis Office, Moscow, ID. 2-Disc CD set (issued October 2001).

Su, S., Xiao, R., Jiang, Z., & Zhang, Y. (2012). Characterizing landscape pattern and ecosystem service value changes for urbanization impacts at an eco-regional scale. *Applied Geography*, *34*(2), 295–305. https://doi.org/10.1016/j.apgeog.2011.12.001

Tobler, W. R. (1979). Smooth Pycnophylactic Interpolation for Geographical Regions. *Journal of the American Statistical Association, 74*(367), 519–530. https://doi.org/10.1080/01621459.1979.10481647

Turner II B. L. T., Kasperson, R. E., Matson, P. A., Mccarthy, J. J., Corell, R. W., Christensen, L., … & Schiller, A. (2003). A framework for vulnerability analysis in sustainability science. *PNASS, 100*(14), 8074–8079.

USACE (2017). US Army Corps of Engineers, Navigation Mission, Controlling Depths Reports, http://www.nan.usace.army.mil/Missions/Navigation/Federal-Operations-Maintenance-Public-Notices/

USCB (1994). Geography Division New York State. TIGER/Line Files. Updates to 1990, U.S. Census Bureau, Washington, DC.

USCB (2000). Census 2000 Summary File 3A, New York, U.S. Census Bureau, Washington, DC.

USFWS (2000). U. S. Fish and Wildlife Service. National Wetlands Inventory [Database], Publication date varies (found in metadata). National Wetlands Inventory website. U.S. Department of the Interior, Fish and Wildlife Service, Washington, D.C. http://www.fws.gov/wetlands/

USGS (1997). *National Elevation Dataset (NED*), US Geological Survey, https://nationalmap.gov/elevation.html

USGS (2014). *NLCD 2001*, USGS National Land Cover Dataset Downloadable Data Collection, January, 2014, https://viewer.nationalmap.gov

Veldkamp, a. & Verburg, P. H. (2004). Modelling land use change and environmental impact. *Journal of Environmental Management*, *72*(1–2), 1–3. https://doi.org/10.1016/j.jenvman.2004.04.004

Wong, D. W. S. (2004). The Modifiable Areal Unit Problem (MAUP). *WorldMinds: Geographical Perspectives on 100 Problems*, (1934), 571–575. https://doi.org/10.1007/978-1-4020-2352-1_93

Zoraghein, H., Leyk, S., Ruther, M., & Buttenfield, B. P. (2016). Exploiting temporal information in parcel data to refine small area population estimates. *Computers, Environment and Urban Systems*, *58*, 19–28. https://doi.org/10.1016/j.compenvurbsys.2016.03.004

CHAPTER 5:

SUMMATION

Overview

At the beginning of this process, three goals guided model development: 1) Utility at the regional (multi-county) scale to provide a coarse filter for the identification of areas with the potential to exhibit sizable decadal residential growth so that such areas could be used to identify biodiversity areas at risk; 2) Reduce impediments to the model's adoption and implementation through the use of existing publicly available digital data; and 3) Provide for model portability across the United States.

In Chapter 2, the Census-based Residential Growth Model for Habitat Vulnerability Assessment, four of the five variables needed are available in digital format. The fifth variable, Birch's Neighborhood Stages (BNS), requires localized building construction permit data to calculate (Birch, 1971). It was a compromise born of necessity due to the lack of statistical significance exhibited by any of the socioeconomic variables comprising the Social Ecology component.

Five independent variables ultimately used in my model are the neighborhood stage of development, the number of housing units built in the prior decade, the regional labor market area, the density of the local road network, and proximity to centers of population. The resulting binomial model predicted each BG's share of new housing units built between the years1990 and 2000.

The model identified 77 of the 2,212 total BG in the study area to be prime candidates for a substantial percentage of the predicted new residential growth

between 2000 and 2010. Of the ten counties in the study area, the top four counties to receive the largest share of new housing were, in order, Albany, Orange, Dutchess and Rensselaer Counties.

Chapter 2 also addressed the mapping of vertebrate species richness by Census BGs and their intersection with the 77 BGs receiving the most shares of anticipated residential housing unit growth. Species mapping demonstrated the following distributions. For amphibians, there was a high number of expected species across all of the region with the lower quartile of species associated with Block Groups near population centers.

There are a large number of expected reptile and amphibian species across the study area with the first quartile of species associated with Block Groups near population centers. However, reptile concentrations also fall into the first quartile for larger non-population center BGs. Across the whole study area, both amphibian and reptile distribution regions envelope the Hudson River area. However, expected reptile distribution areas are more fragmented than expected amphibian distribution regions.

Mammal species distributions are heavily affected by BG population and metropolitan centers, with the number of species increasing inversely with distance. The regions of expected mammal distribution are consolidated blocks exhibiting little fragmentation.

Expected breeding bird species distributions exhibit an inverse relationship especially noticeable around the metropolitan areas. In higher populated areas, the bird species concentration increased while the rest of the study area displayed no apparent pattern.

Concentrations of species classified as TES (threatened, endangered, or of special concern) which rank in the third and fourth quartiles of all such BGs, comprise most of the study. The lower two quartiles of BGs occupy much smaller areas concentrated in the urban areas.

The results from this model support the hypothesis that residential growth can be modeled using Census data and that the results from such modeling can efficiently identify biodiversity risk areas, thereby enabling the targeting of education and mitigation efforts.

Modeling Residential Growth and Habitat Vulnerability on a Regional Scale, Chapter 3, addresses the replacement of the BNS variable and inclusion of the Census 2010 data for model selection and validation. This chapter focused on the hypothesis that modeling the spatial distribution of residential housing growth with Census Block Group (BG) level data can be accomplished without requiring non-digital local data, specifically the data for the calculation of the BNS variable while providing similar predictive power for a suitable coarse filter biodiversity risk application.

The resulting model, which explained 38% of the variance, identified the BGs with residential growth ranking them in the top 5% of all BGs producing a 43% match with the 106 BGs within the top 5% of high growth BGs as determined by actual Census 2010 data. An error matrix of the model results produced a true positive rate of 40.5% and an overall accuracy rate of 94.0%.

When compared to the first model, the 77 top growing BGs of this model matched 33 of the actual top BGs exhibiting growth whereas the previous model only matched

3 of the high growth BGs identified in the 2010 Census. This result demonstrates a substantial improvement in spatial accuracy.

This spatial accuracy is essential when seeking to identify biodiversity areas at risk. The intersection of biodiversity data with the model output resulted in the identification of eight BGs of the 211 biodiversity areas as coincident with high growth areas. According to the Census 2010 data, only five of the high biodiversity BGs were also BGs with substantial residential housing growth, and of those five, the model correctly identified two. The model has a 25% accuracy rate in predicting which high biodiversity area are at risk from high residential growth.

The success of this effort supported the hypothesis that local non-digital data were not necessary to maintain the predictability of growth areas and the applicability of the model to biodiversity conservation at the regional scale.

In an attempt to improve the accuracy of the prediction of residential housing unit increases, an investigation of dasymetric mapping was undertaken, as presented in Chapter 4, Impact of Dasymetric Mapping on a Census-based Regional Scale Residential Growth Model. I hypothesized, based on the earlier model's success, that the spatial accuracy of the existing residential growth model can be improved using dasymetric methods to distribute BG data to spatial subsets while maintaining the effectiveness of the model to identify at-risk biodiversity areas. As the chapter explains, dasymetric mapping was not successful in increasing model accuracy.

The NLCD2001 Land Cover dataset (USGS, 2014;Homer et al., 2007) was acquired to accomplish the dasymetric mapping of the Census 2000 BG variables to a more higher spatial resolution. The 15 NLCD 2001 classes found in New York State

underwent an aggregation into five: rural, low-density residential, medium-density residential, high-density residential/commercial/industrial, and non-developable.

Because the identification of sizable areas prone to residential growth that might pose a threat to vertebrate species habitats was the objective, the quarter-acre cell size represented by the 30-meter cell resolution is unnecessarily detailed. Resampling the NLCD2001 layer to a 180-meter cell raster provides a 3.25 hectare resolution.

The generation of a non-developable layer which includes surface hydrography, federally regulated wetlands, slopes in excess of 25%, and areas at or above a NY-GAP stewardship level of 3, provides a means to add detail to the NLCD's non-developable class.

The NLCD2001 layer was intersected with the Census 2000 BGs to create more than14,000 BG land cover polygons. A BG's total housing units were proportionally allocated to each of the BG's land cover subset. The other model factors for the BG subset polygons involved calculation of a like manner.

Weak results from the model, when applied to this dataset, prompted an investigation into the cause. There were two issues identified: the confusion of housing density areas in the NLCD classification and misalignment of Census 2000 BG boundaries compared to the NLCD 2001 layer and other authoritative layers.

Another land cover layer for the study area did not exist, and the creation of one would interfere with a primary goal, as would "re-drawing" the BG boundaries.

The NHGIS datasets which used dasymetric mapping to relate the Census 2000 BG data to Census 2010 BG boundaries offered an alternative (Manson et al., 2017). The NHGIS dataset provides the 2000 and 2010 decadal surveys in the Census 2010 BG geography which is of a higher spatial accuracy.

Model variables were re-calculated using the NHGIS Census 2010 BG data values for 2000 and 2010. Using these values and the NHGIS Census 2010 BG boundaries, new NLCD 2001 land cover BG subsets were generated and populated with values. Another run of the model again produced poor results.

Applying the early model version using full BG data from the NHGIS Census 2010 tested whether the NLCD 2001 dataset was responsible for the model results. The improvement from removing the NLCD 2001 factor was minimal, and the results were still substantially weaker than using the GeoLytics 2000 dataset. A GeoLytics 2010 data set was acquired to evaluate if the NHGIS Census 2010 data set may be causing the discrepancy (GeoLytics, 2010). The GeoLytics 2010 data set is generated using traditional Census relationship files to translate the Census 2000 BG data to Census 2010 BG boundaries. Model runs on the GeoLytics 2010 data set showed no improvement.

The results presented in Chapter 4 supported the null hypothesis; the introduction of dasymetric mapping techniques did not improve the existing model but instead altered the model in such a way as to make it unacceptable for growth predictions and subsequent detection of at-risk biodiversity areas.

Discussion

Efforts to develop a coarse filter model for the prediction of residential growth that has applicability at a regional scale has produced mixed results. Some of these results produced very positive signs as described in Chapters 2 and 3. These positive results enable the demonstration of the value of a coarse filter predictive model for the identification of biodiversity areas at risk from potential future residential development. The availability of such a model provides a means to target mitigating efforts and thereby increase their efficiency.

The lack of any positive results in the last attempt to refine the model has produced the following conclusions:

First, the translation of the independent variables by both the NHGIS and the GeoLytics methods have altered the relationship previously modeled.

Second, the primary goal of only using publically available digital data may be overly restrictive and thereby fails to include localized factors which may have a substantial impact on development. This lack of localized data may also have negatively impacted the attempt to apply dasymetric mapping to the Block Group data. Certainly, the misalignment of the BG boundaries introduces a serious problem for comparing decadal Census data. This issue will likely resolve itself over the years as more accurate initial boundary data are collected and mapping errors reduced.

Third, the collection of land use data used for the dasymetric mapping effort was ineffective in rural and low-density housing areas due to the many inclusions of parks, golf courses, and cemeteries. It is possible that a land use mapping effort undertaken with land use categories specifically designed to delineate the rural, low density, moderate density, and high-density residential areas may have enabled the success of the dasymetric mapping effort. The success of this effort is especially probable if coupled with the realignment of the incorrect BG boundaries. Again, perhaps the restriction of “available” data may have been too limiting.

Independent of the lack of improvement in the model’s predictive power from the implementation of dasymetric mapping, there is an argument to be made that using current conditions and past development patterns can only approximate development trends. Many attempting to model residential growth believe any attempt to predict residential growth must include factors preceding the home buyer’s decisions and may be independent of the buyer’s characteristics. Most substantial housing increases are likely the result of government regulations and commercially developed subdivisions,

not individuals purchasing unimproved land and then building their home. The governmental agency charged with local land use control makes decisions in the planning process that influence the amount and location of the available developable land, such as zoning-based restrictions and utility expansions. Private and commercial developers make decisions involving land speculation based on the actions of the governmental authorities. These decisions may range from simple land purchases to actual land improvements and perhaps even the construction of housing units. The developer is the decision maker who moves the land from unimproved to developable, with the help of the political system which expands or denies services such as roads, water, sewer and other utilities. I successfully developed a model useful at a regional scale based on the Census BG geography, but it would seem that increasing the model's spatial resolution may not be possible without incorporating more localized data.

References

Birch, D.L. (1971). Toward a Stage Theory of Urban Growth. *Journal of American Institute of Planners*, 37(2), 78-87.

GeoLytics (2010). Census Data 2000 in 2010 Boundaries, GeoLytics, Inc., Retrieved from http://www.geolytics.com/USCensus,Normalized-Data,Categories.asp

Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C.,…Wickham, J. (2007). Completion of the 2001 National Land Cover Database for the Conterminous United States. *Photogrammetric Engineering and Remote Sensing*, 73(4), 337-341.

Manson S., Schroeder J., Van Riper D., & Ruggles S. (2017). *IPUMS National Historical Geographic Information System: Version 12.0 [Database]*. Minneapolis, MN: University of Minnesota. http://doi.org/10.18128/D050.V12.0

USGS (2014). *NLCD 2001*, USGS National Land Cover Dataset Downloadable Data Collection, January, 2014, https://viewer.nationalmap.gov

BIBLIOGRAHY

American Society of Planning Officials (1959). *Hillside Development. Planning Advisory Service. Information Rpt.* (No. 126). Chicago, IL: Clark, R.A.

Allen, E. (2001). INDEX: Software for Community Indicators. In R. K. Brail & E. Klosterman (Eds.), *Planning Support Systems* (pp 229-261). Redlands, CA: ESRI Press.

Alonso, W. (1964). *Location & Land Use - Toward a General Theory of Land Rent*. Cambridge, MA: Harvard Press.

Berke, P.R., Godschlk, D.R., Kaiser, E.J. & Rodriguez, D.A. (2006). *Urban Land Use Planning*, 5th Ed., Urbana, IL: University of Illinois Press.

Birch, D.L. (1971). Toward a Stage Theory of Urban Growth. *Journal of American Institute of Planners*, 37(2), 78-87.

Britannica (2016). In *Encyclopedia Britannica* online. Retrieved from Encyclopedia https://www.britannica.com/science/Newtons-law-of-gravitation

Burchell, R. W. & Mukherji, S. (2003). Conventional Development Versus Managed Growth: The Costs of Sprawl. *American Journal of Public Health*, 93, 1534-1540.

Campbell, J.B. & Wynne, R.H. (2011). *Introduction to Remote Sensing*. New York, NY: The Guilford Press.

Clearwater (2017). Hudson River Clearwater Sloop. Retrieved from http://www.clearwater.org/

Congalton, R.G. & Green, K. (1999). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, New York, NY: Lewis Publishers.

Dark, S. J., & Bram, D. (2007). The modifiable areal unit problem (MAUP) in physical geography. *Progress in Physical Geography*, *31*(5), 471–479. https://doi.org/10.1177/0309133307083294

Edmonston, B. (1975). *Population Distribution in American Cities*. Lexington, CT: Lexington Books.

Eicher, C.L. & Brewer, C.A. (2001). Dasymetric mapping and areal interpolation: implementation and evaluation, *Cartography and Geographic Information Science*, 28(2), 125-138.

Foot, D. (1981). *Operational Urban Models an Introduction*. Bungay, UK: Chaucer Press.

GeoLytics (2000a). *Census 2010 Summary File 1 in 2000 Boundaries*, Geolytics CD-ROM.

GeoLytics (2000b). *Census 1990 Summary File 1 in 2000 Boundaries*, Geolytics CD-ROM.

GeoLytics (2010). Census Data 2000 in 2010 Boundaries, GeoLytics, Inc., Retrieved from http://www.geolytics.com/USCensus,Normalized-Data,Categories.asp

Gustafson, E.J., Hammer, R.B., Radeloff, V.C. & Potts, R.S. (2005). The Relationship between Environmental Amenities and Changing Human Settlement Patterns between 1980 and 2000 in the Midwestern USA. *Landscape Ecology*, 20, 773–789.

Hammer, R.B., Stewart, S.I., Winkler, R.L. et al. (2004). Characterizing dynamic spatial and temporal residential density patterns from 1940-1990 across the North Central United States. *Landscape and Urban Planning*, 69, 183-199.

Hawbaker, T.J., Radeloff V.C., Hammer R.B., & Clayton M.K. (2005). Road density and landscape pattern in relation to housing density, and ownership, land cover, and soils. *Landscape Ecology*, 20, 609–625.

Holt, J.B., Lo, C.P., & Hodler, T.W. (2004). Dasymetric Estimation of Population Density and Areal Interpolation of Census Data. *Cartography and Geographic Information Science*, 31(2), 103–121. https://doi.org/10.1559/1523040041649407

Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C.,…Wickham, J. (2007). Completion of the 2001 National Land Cover Database for the Conterminous United States. *Photogrammetric Engineering and Remote Sensing*, 73(4), 337-341.

Hoover E. M. & Vernon R. (1959). *Anatomy of a Metropolis: the changing distribution of people and jobs in the New York Metropolitan Region*. Cambridge, MA: Harvard University Press.

Hu, W. (2000, June 7). Industrial Growth Threatens Scenic Hudson River Valley, Group Warns. *New York Times*. Retrieved from http://www.nytimes.com /2000/06/27/ nyregion/ industrial-growth-threatens-scenic-hudson-river-valley-group-warns.html

Hudson River Estuary Program (1987). Hudson River Estuary Management Act. *NYS Environmental Conservation Law*, 11-0306. Retrieved from http://www.dec.ny.gov/lands/43251.html

Jia, P., & Gaughan, A. E. (2016). Dasymetric modeling: A hybrid approach using land cover and tax parcel data for mapping population in Alachua County, Florida. *Applied Geography*, *66*, 100–108. https://doi.org/10.1016/j.apgeog.2015.11.006

Kim, H., & Yao, X. (2010). Pycnophylactic interpolation revisited: Integration with the dasymetric-mapping method. *International Journal of Remote Sensing, 31*(21), 5657–5671. https://doi.org/10.1080/01431161.2010.496805

Klosterman, R.E. (2008). A New Tool for a New Planning: The What if? Planning Support System. In R.K. Brail (Ed), *Planning Support Systems for Cities and Regions*, (pp. 85-99). Cambridge, MA: Lincoln Institute of Land Policy.

Kwartler, M. & Bernard, R.N. (2001). CommunityViz: An Integrated Planning Support System. In R. K. Brail & Klosterman, E. (Eds.), *Planning Support Systems*, (pp. 285-308). Redlands, CA: ESRI Press.

Langford, M. (2007). Rapid facilitation of dasymetric-based population interpolation by means of raster pixel maps. *Computers, Environment and Urban Systems, 31*(1), 19–32. https://doi.org/10.1016/j.compenvurbsys.2005.07.005

Landis, J. & Zhang M. (1993). The second generation of the California urban futures model. Part 2: Specification and calibration results of the land-use change submodel. In *Environment and Planning, B. Planning and Design*, 25(6), 795-824, Bristol, UK:Pion Limited.

Lillesand, T.M., & Kiefer R.W. (2000). *Remote Sensing and Image Interpretation*, New Jersey: John Wiley Press.

Lowery, I.S. (1964). *A Model of Metropolis, Memorandum RM-4035-RC*, Santa Monica, CA: The Rand Corporation.

Manson S., Schroeder J., Van Riper D., & Ruggles S. (2017). *IPUMS National Historical Geographic Information System: Version 12.0 [Database]*. Minneapolis, MN: University of Minnesota. http://doi.org/10.18128/D050.V12.0

Mennis, J. (2003). Generating Surface Models of Population Using Dasymetric Mapping. *Professional Geographer*. https://doi.org/10.1111/0033-0124.10042

Mennis, J., & Hultgren, T. (2006). "Intelligent" Dasymetric Mapping and Its Comparison to Other Areal Interpolation Techniques. *Cartogis.Org,* (215), 1–9. Retrieved from http://www.cartogis.org/publications/autocarto-2006/mennishultgren.pdf

Metropolitan Museum of Art (2017). *Heilbrunn Timeline of Art History*. Retrieved from https://www.metmuseum.org/toah/hd/hurs/hd_hurs.htm2

Mitsova, D., Esnard, A. M., & Li, Y. (2012). Using enhanced dasymetric mapping techniques to improve the spatial accuracy of sea level rise vulnerability assessments. *Journal of Coastal Conservation*, *16*(3), 355–372. https://doi.org/10.1007/s11852-012-0206-3

Moudon, A. & Hubner, M. (2000). Elements of a General Framework for Land Supply and Capacity Monitoring. In A. Moudon & M. Hubner (Eds), *Monitoring land supply with geographic information systems: Theory, practice, and parcel-based approaches* (pp. 41-64). New York, NY: John Wiley & Sons.

NYSDEC (1999) *Environmental Conservation Law of New York, Section 11-0535 and 6 NYCRR (New York Code of Rules and Regulations) Part 182*. Albany, NY: New York State Department of Environmental Conservation.

NYSDEC (2013). *New York State Regulatory Freshwater Wetlands, 1999-2013*. Albany, NY: NYS Dept. of Environmental Conservation. http://cugir.mannlib.cornell.edu/READMES/fwrdme.html

NYSDOS (2001). *Adopting Zoning for the First Time*, Local Government Technical Series, James A. Coon, Albany, NY: New York State Department of State.

Ottensmann, J.R. (1977). Urban Sprawl, Land Values and the Density of Development, *Land Economics*, 53(4) 389-400.

Perkins, G.C, Kutt, A.S, Vanderduys E.P, & Perry J.J. (2013). Evaluating the costs and sampling adequacy of a vertebrate monitoring program. *Aust. Zool.*, 36, 373–380.

Riverkeeper (2017). https://www.riverkeeper.org/riverkeeper-mission/our-story/

Sanchez, T. W. (2000). Land use and growth impacts from highway capacity increases. *Journal of Urban Planning and Development*, 130(2), 75–82. https://doi.org/10.1061/(ASCE)0733-9488(2004)130:2(75)

Sayweh, P.W. (1998). *Combining a Growth Allocation Model and GIS Results to Assess Suitable Biodiversity Management Areas: A Prototype for Hudson Valley Biodiversity Program in New York State* (Unpublished MRP thesis), Cornell University Graduate School, Ithaca, NY.

Scenic Hudson (2017). https://www.scenichudson.org/about/history

Schroeder, J. P. (2017). Hybrid areal interpolation of census counts from 2000 blocks to 2010 geographies. *Computers, Environment and Urban Systems* 62, 53-63. http://dx.doi.org/10.1016/j.compenvurbsys.2016.10.001

Scott, J.M., Davis F., Csuti B., Noss R., Butterfield B., Groves C.,…Wright R.G. (1993). *Gap Analysis: A Geographic Approach to Protection of Biological Diversity*. Wildlife Monographs 123:1-41.

Sleeter, R., & Gould, M. (2007). Geographic information system software to remodel population data using dasymetric mapping methods. *US Geological Survey, Techniques and Methods*, 11-C2. Reston, VA: USGS. Retrieved from http://pubs.usgs.gov/tm/tm11c2/

Smith, C.R., DeGloria S.D., Richmond M.E., Gregory S.K., Laba M., Smith S.D.,…Weber J.T. (2001a). *A Gap Analysis of New York NY-GAP. Final Report and Data*. U.S. Geological Survey, Biological Resources Division, National Gap Analysis Office, Moscow, ID. 2-Disc CD set (issued October 2001).

Smith, C.R., DeGloria S.D., Richmond M.E., Gregory S.K., Laba M., Smith S.D.,… Hill E.A. (2001b). *An Application of Gap Analysis Procedures to Facilitate Planning for Biodiversity Conservation in the Hudson River Valley, Final Report, Part 1: Gap Analysis of the Hudson River Valley and Part 2: Atlas of Predicted Ranges for Terrestrial Vertebrates in the Hudson River Valley*. New York Cooperative Fish and Wildlife Research Unit. Ithaca, NY:Department of Natural Resources, Cornell University.

Smith S.D., Brown W.A., Smith C.R., & Richmond M.E. (2004). *Habitat Vulnerability Assessment for the Hudson River Valley: Final Report.* New York Cooperative Fish and Wildlife Research Unit. Ithaca, NY:Department of Natural Resources, Cornell University.

Smith S.D., Brown W.A., Smith C.R., & Richmond M.E. (2005). *Habitat Vulnerability Assessment in the Hudson River Valley*, GAP Analysis Bulletin No. 13, December, 37-43.

Su, S., Xiao, R., Jiang, Z., & Zhang, Y. (2012). Characterizing landscape pattern and ecosystem service value changes for urbanization impacts at an eco-regional scale. *Applied Geography*, *34*(2), 295–305. https://doi.org/10.1016/j.apgeog.2011.12.001

Theobald, D.M., Hobbs, N.T., Bearly, T. et al. (2000). Incorporating Biological Information in Local Land-Use Decision Making: Designing a System for Conservation Planning, *Landscape Ecology*, 15:35-45.

Theobald, D.M. (2001). Land-Use Dynamics Beyond the American Urban Fringe. *The Geographical Review*, 91 (3): 544-564.

Theobald, D. M., & Hobbs, N. T. (2002). A framework for evaluating land use planning alternatives: Protecting biodiversity on private land. *Ecology and Society*, *6*(1).

Tobler, W. R. (1979). Smooth Pycnophylactic Interpolation for Geographical Regions. *Journal of the American Statistical Association*, *74*(367), 519–530. https://doi.org/10.1080/01621459.1979.10481647

Tolbert, C. M. & Killian M. S. (1987). *Labor Market Areas for the United States*. Staff Report No. AGES870721. Washington, DC: Economic Research Service, USDA.

Tolbert C. M. & Sizer M. (1996). *U.S. Commuting Zones and Labor Market Areas: a 1990 Update*. U.S. Department of Agriculture Economic Research Service Staff Paper No. 9614. September, Washington, DC: Economic Research Service, USDA.

Turner II B. L. T., Kasperson, R. E., Matson, P. A., Mccarthy, J. J., Corell, R. W., Christensen, L., … & Schiller, A. (2003). A framework for vulnerability analysis in sustainability science. *PNASS*, *100*(14), 8074–8079.

USACE (2017). US Army Corps of Engineers, Navigation Mission, Controlling Depths Reports, http://www.nan.usace.army.mil/Missions/Navigation/Federal-Operations-Maintenance-Public-Notices/

USCB (1994). Geography Division New York State. TIGER/Line Files. Updates to 1990, U.S. Census Bureau, Washington, DC.

USCB (1999). Housing Unit Review File. New York State. Population Estimates Branch, U.S. Census Bureau, Washington, DC.

USCB (2000). Census 2000 Summary File 3A, New York, U.S. Census Bureau, Washington, DC.

USCB (2002). Census 2000 Summary File 3. Technical Documentation. U.S. Census Bureau, Washington, DC.

USFWS (2000). U. S. Fish and Wildlife Service. National Wetlands Inventory [Database], Publication date varies (found in metadata). National Wetlands Inventory website. U.S. Department of the Interior, Fish and Wildlife Service, Washington, D.C. http://www.fws.gov/wetlands/

USGS (1997). *National Elevation Dataset (NED*), US Geological Survey, https://nationalmap.gov/elevation.html

USGS (2014). *NLCD 2001*, USGS National Land Cover Dataset Downloadable Data Collection, January, 2014, https://viewer.nationalmap.gov

Veldkamp, A, & Verburg, P. H. (2004). Modelling land use change and environmental impact. *Journal of Environmental Management*, *72*(1–2), 1–3. https://doi.org/10.1016/j.jenvman.2004.04.004

White, D., Minotti, P.G., Barczak, M.J. et al. (1997). Assessing Risks to Biodiversity from Future Landscape Change, *Conservation Biology,* 11(2), April 1997, 349-360.

Wong, D. W. S. (2004). The Modifiable Areal Unit Problem (MAUP). *WorldMinds: Geographical Perspectives on 100 Problems*, (1934), 571–575. https://doi.org/10.1007/978-1-4020-2352-1_93

Zoraghein, H., Leyk, S., Ruther, M., & Buttenfield, B. P. (2016). Exploiting temporal information in parcel data to refine small area population estimates. *Computers, Environment and Urban Systems*, *58*, 19–28. https://doi.org/10.1016/j.compenvurbsys.2016.03.004

Yang, Y., Brown W., & Smith S. (2001). *Using Neighborhood Growth Stage Model to Analyze New Residential Development in Ten Counties in Hudson River Valley Area, New York State.* Population & Development Program Working Papers. Ithaca, NY: Program in Applied Demographics, Cornell University.